

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 11:01:19 : Search time 31 Seconds  
(without alignments)  
272.514 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41

Sequence: 1 TPLGPKWPEVFGRLASPGF.....GEYANDQERRWTLTAPPGYR 41

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Database: 671580 seqs, 206047115 residues

Size: 16

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database:

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_rv1rus:\*  
16: SP\_bacteriap:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	175	4 Q9UMV3	Q9umv3 homo sapien
2	41	100.0	181	4 Q9ULC7	Q9ulc7 homo sapien
3	41	100.0	185	4 Q9UBP3	Q9ubp3 homo sapien
4	16	39.0	16	4 Q9UC48	Q9uc48 homo sapien

#### ALIGNMENTS

RESULT 1  
Q9UMV3  
AC Q9UMV3: PRELIMINARY: PRT: 175 AA.  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

or 102 (b) prior and  
to oligo class

DE Mannose binding lectin-associated serine protease-2 related protein,  
DE MAP19 (19kDa) precursor (Fragment).  
GN MASP-2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER.  
RX MEDLINE=99192764; PubMed=10092804;  
RA Slover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,  
RA Jensenius J.C., Schwaeble W.J.;  
RT "Two constituents of the initiation complex of the mannose-binding  
RT lectin activation pathway of complement are encoded by a single  
RT structural gene.";  
RL J. Immunol. 162:3481-3490(1999).  
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
DR EMBL: Y18282; CAB50729.1; -  
DR HSSP: P00736; IAPQ.  
DR MEROPS: S01.229; -  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000859; CUB\_domain.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR Pfam: PF00431; CUB; 1.  
DR SMART: SM00042; CUB; 1.  
DR SMART: SM00179; EGF\_Ca; 1.  
DR PROSITE: PS00101; ASX\_HYDROXYL; UNKNOWN\_1.  
DR PROSITE: PS01180; CUB; 1.  
DR PROSITE: PS01186; EGF\_1.  
DR PROSITE: PS01187; EGF\_Ca; 1.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease;  
KW Repeat; Signal.  
FT NON\_TER 1 1  
FT SIGNAL <1 5  
FT CHAIN 6 175  
FT FT  
SQ SEQUENCE 175 AA; 19504 MW; 474FB56221CD28D1 CRC64;  
Query Match 100.0%; Score 41; DB 4; Length 175;  
Best Local Similarity 100.0%; Pred. No. 3.8e-35;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TPLGPKWPEVFGRLASPGFGEYANDQERRWTLTAPPGYR 41  
Db 6 TPLGPKWPEVFGRLASPGFGEYANDQERRWTLTAPPGYR 46  
RESULT 2  
Q9ULC7  
ID Q9ULC7: PRELIMINARY: PRT: 181 AA.  
AC Q9ULC7:  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE MBL-associated serine protease(MASP)-2 (Fragment).  
GN MASP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Takahashi M., Fujita T.;  
RT "Partial genomic structure of human MBL-associated serine protease  
RT (MASP)-2 (from exon 1 to exon 5).";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
DR EMBL: AB033742; BAA85659.1; -  
DR HSSP: P00736; IAPQ.  
DR MEROPS: S01.229; -

```

DR InterPro: IPR000152; Asx_hydroxy1.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Protease; Repeat.
FT NON_TER 181
SO SEQUENCE 181 AA; 20171 MW; 5FB229C21B89DDBC CRC64;

Query Match 100.0%; Score 41; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 3.9e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TPLGPKWPEPVGRGLASPGFCEVANDDERRWTLTAPPCXR 41
|||||
16 TPLGPKWPEVPVGRGLASPGFEVANDDERRWTLTAPPGR 56

RESULT 3
Q9UBP3 PRELIMINARY: PRT; 185 AA.
ID Q9UBP3
AC Q9UBP3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Manose binding lectin-associated serine protease-2 related protein,
DE MAR19 (19 kDa) precursor (19 kDa) (MBL-associated protein Map19).
GN MASP-2 OR SAMP OR MASP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA MEDLINE=99192764; PubMed=10092804;
RA Stover C.M., Thiel S., Thelem W., Lynch N.J., Vorp-Jensen T.,
RA Jensenius J.C., Schwaible W.J.;
RT "Two constituents of the initiation complex of the mannose-binding
RT lectin activation pathway of complement are encoded by a single
RT structural gene.";
RT J. Immunol. 162:3481-3490(1999).
2]
RP SEQUENCE FROM N.A.
RC MEDLINE=99262288; PubMed=10330290;
RA Takahashi M., Fujita T.;
RA Takahashi M., Endo Y., Fujita T.;
RT "A truncated form of mannose-binding lectin-associated serine
RT protease (MASP)-2 expressed by alternative polyadenylation is a
RT component of the lectin complement pathway.";
RT Int. Immunol. 11:859-863(1999).
[3]
RP SEQUENCE FROM N.A.
RC RA Takahashi M., Fujita T.;
RT "Partial genomic structure of human MBL-associated serine protease
RT (MASP)-2 (from exon 1 to exon 5).";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC Park D., Kim B., Baek K., Yoon J.;
RT "Structure of Human MASP-2 Gene.";
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: Y18284; CAB50731.1; -.
DR EMBL: AB008047; BAA78616.1; -.
DR EMBL: AB033742; BAAB5658.1; -.
DR EMBL: Y18281; CAB50728.1; -.
DR EMBL: Y18283; CAB50730.1; -.

```

[illegible]

```
Query March      39.0%; Score 16; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. NO. 6.7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY    15 L A S P G F P G E Y A N D O E R   30
      ||| | | | | | | | | | | | | | | |
Db     1 L A S P G F P G E Y A N D O E R   16
```

Search completed: January 11, 2003, 11:05:43  
Job time : 31 secs

Search completed: January 11, 2003, 11:05:43  
Job time : 31 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 11:05:49 : Search time 11 Seconds  
(without alignments)  
72.313 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41  
Sequence: 1 TPLGPKWPEPVFGRLASPGF.....GEYANDQERRWTLTAPPGR 41

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Size: 16  
118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: listing first 500 summaries

Database: Published Applications-AA:\*

1: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	41	US-09-874-198-1	Sequence 1, Appli
2	41	100.0	41	US-09-874-238-1	Sequence 1, Appli
3	41	100.0	686	US-09-874-198-2	Sequence 2, Appli
4	41	100.0	686	US-09-874-238-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-874-198-1  
; Sequence 1, Application US/09874198  
; Patent No. US20020082208A1  
; GENERAL INFORMATION:  
; APPLICANT: Jensenius, Jens Chr.  
; APPLICANT: Thiel, Steffen  
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND  
; FILE REFERENCE: 09011-002002

; CURRENT APPLICATION NUMBER: US/09/874,198  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: 09/054,218  
; PRIOR FILING DATE: 1998-04-02  
; PRIOR APPLICATION NUMBER: 60/042,678  
; PRIOR FILING DATE: 1997-04-03  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 41  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-874-198-1

Query Match 100.0%; Score 41; DB 10; Length 41;  
Best Local Similarity 100.0%; Pred. No. 3.6e-35;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGR 41  
|||||  
Db 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGR 41

RESULT 2  
US-09-874-238-1  
; Sequence 1, Application US/09874238  
; Patent No. US20020082209A1  
; GENERAL INFORMATION:  
; APPLICANT: Jensenius, Jens Chr.  
; APPLICANT: Thiel, Steffen  
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND  
; FILE REFERENCE: 09011-002003  
; CURRENT APPLICATION NUMBER: US/09/874,238  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: 09/054,218  
; PRIOR FILING DATE: 1998-04-02  
; PRIOR APPLICATION NUMBER: 60/042,678  
; PRIOR FILING DATE: 1997-04-03  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 41  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-874-238-1

Query Match 100.0%; Score 41; DB 10; Length 41;  
Best Local Similarity 100.0%; Pred. No. 3.6e-35;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGR 41  
|||||  
Db 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGR 41

RESULT 3  
US-09-874-198-2  
; Sequence 2, Application US/09874198  
; Patent No. US20020082208A1  
; GENERAL INFORMATION:  
; APPLICANT: Jensenius, Jens Chr.  
; APPLICANT: Thiel, Steffen  
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND  
; FILE REFERENCE: 09011-002002  
; CURRENT APPLICATION NUMBER: US/09/874,198  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: 09/054,218  
; PRIOR FILING DATE: 1998-04-02  
; PRIOR APPLICATION NUMBER: 60/042,678  
; PRIOR FILING DATE: 1997-04-03  
; NUMBER OF SEQ ID NOS: 8





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 11:03:54 : Search time 17 Seconds  
(without alignments)  
70.961 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41  
Sequence: 1 TPLGFKWPEPVFGRLASPGF.....GEVANDQERRMTLTAPGVR 41

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

sd: 262574 seqs, 29422922 residues

lze: 16

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database: Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/CTUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Score	Match length	ID	Description
Matches found				

Search completed: January 11, 2003, 11:06:46  
Job time: 17 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:54:24 ; Search time 10 seconds  
(without alignments)  
170.053 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41

Sequence: 1 TPLGPKWPEVYGRLASPGF.....GEVANDQERWTLTPPGYR 41

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Database: 112892 seqs, 41476328 residues

Size: 16

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : SwissProt\_40:\*

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	686	1 MAS2_HUMAN	000187 homo sapien

# ALIGNMENTS

Result 1  
MAS2\_HUMAN STANDARD; PRT; 686 AA.  
000187: 075754; 091270; 098280;  
30-MAY-2000 (Rel. 39, last sequence update)  
15-JUN-2002 (Rel. 41, last annotation update)  
DE Mannan-binding lectin serine protease 2 precursor (EC 3.4.21.-)  
DE (Mannan-binding protein associated serine protease 2) (MASP-2)  
DE (MBL-associated serine protease 2).  
GN MASP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=97242412; PubMed=9087411;  
RA Thiel S., Jensen T.V., Stover C.M., Schwaebel W.J., Laursen S.B.,  
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,  
RA Reid K.B.M., Jensenius J.C.;  
RT "A second serine protease associated with mannan-binding lectin that  
RT activates complement."  
RL Nature 386:506-510(1997).  
RN [2]  
RP SEQUENCE FROM N.A.

RA Thiel S., Vorup-Jensen T., Stover C.M., Schwaebel W.J., Laursen S.B.,  
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,  
RA Reid K.B.M., Jensenius J.C.;  
RT "Identification and characterization of a novel protein of the human  
RT complement system, mannan-binding lectin-associated serine protease-2  
RT (MASP-2)."  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99192764; PubMed=10092804;  
RA Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,  
RA Jensenius J.C., Schwaebel W.J.;  
RT "Two constituents of the initiation complex of the mannan-binding  
RT lectin activation pathway of complement are encoded by a single  
RT structural gene".  
RL J. Immunol. 162:3481-3490(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Park D., Kim B., Baek K., Yoon J.;  
RT "Structure of human MASP-2 gene."  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: TRYPSIN PROTEASE THAT PRESUMABLY PLAYS AN IMPORTANT ROLE  
CC IN THE INITIATION OF THE MBL COMPLEMENT ACTIVATION PATHWAY. AFTER  
CC ACTIVATION IT CLEAVES C4 GENERATING C4A AND C4B.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -!- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: Y09926; CAA71059.1; -;  
DR EMBL: X98400; CAA67050.1; -;  
DR EMBL: Y18287; CAB50735.1; -;  
DR EMBL: Y18286; CAB50733.1; -;  
DR EMBL: AF321562; AAG50274.1; -;  
DR EMBL: AF321558; AAG50274.1; JOINED.  
DR EMBL: AF321559; AAG50274.1; JOINED.  
DR EMBL: AF321560; AAG50274.1; JOINED.  
DR EMBL: AF321561; AAG50274.1; JOINED.  
DR HSSP: P00763; IDPO.  
DR MEROPS: S01.229; -;  
DR GeneW: HGNC:6902; MASP2.  
DR MIM: 605102; -;  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000859; CUB\_domain.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR InterPro: IPR000436; Sush1\_SCR\_CCP.  
DR Pfam: PF00084; sush1; 2.  
DR Pfam: PF00089; trypsin; 1.  
DR Pfam: PF00431; CUB; 2.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00032; CCP; 2.  
DR SMART: SM00042; CUB; 2.  
DR SMART: SM00179; EGF\_CA; 1.  
DR SMART: SM00020; TRYP-Spc; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE: PS01180; CUB; 2.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS01187; EGF\_CA; 1.  
DR PROSITE: PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE: PS00134; TRYPSIN\_HIS; FALSE\_NEG.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Complement pathway; Serine protease; Protease;

KW Glycoprotein; Sushi; Repeat; Signal; EGF-like domain; Hydroxylation.  
FT SIGNAL 1 15 POTENTIAL.  
FT CHAIN 16 686 MANNAN-BINDING LECTIN SERINE PROTEASE 2.  
FT CHAIN 16 444 MANNAN-BINDING LECTIN SERINE PROTEASE 2 A  
FT CHAIN 445 686 CHAIN.  
FT CHAIN MANNAN-BINDING LECTIN SERINE PROTEASE 2 B  
FT DOMAIN 16 137 CUB 1.  
FT DOMAIN 138 181 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 184 296 CUB 2.  
FT DOMAIN 299 362 SUSHI 1.  
FT DOMAIN 365 431 SUSHI 2.  
FT DOMAIN 445 686 SERINE PROTEASE.  
FT ACT\_SITE 483 483 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 532 532 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 633 633 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT MOD\_RES 158 158 HYDROXYLATION (POTENTIAL).  
FT DISULFID 72 90 POTENTIAL.  
FT DISULFID 142 156 POTENTIAL.  
FT DISULFID 152 165 POTENTIAL.  
FT DISULFID 167 180 POTENTIAL.  
FT DISULFID 184 211 POTENTIAL.  
FT DISULFID 241 259 POTENTIAL.  
FT DISULFID 300 348 POTENTIAL.  
FT DISULFID 328 361 POTENTIAL.  
FT DISULFID 366 412 POTENTIAL.  
FT DISULFID 396 430 POTENTIAL.  
FT DISULFID 434 552 INTERCHAIN (POTENTIAL).  
FT DISULFID 598 618 POTENTIAL.  
FT DISULFID 629 660 POTENTIAL.  
FT CONFLICT 361 362 MISSING (IN REF. 3).  
FT CONFLICT 371 371 D -> Y (IN REF. 4).  
FT CONFLICT 372 372 L -> LCS (IN REF. 3).  
FT CONFLICT 442 442 G -> E (IN REF. 4).  
FT CONFLICT 447 447 G -> E (IN REF. 4).  
FT CONFLICT 461 462 MISSING (IN REF. 3).  
FT CONFLICT 473 473 L -> LIL (IN REF. 3).  
SQ SEQUENCE 686 AA; 75685 MM; 4E34DED159448A2A CRC64;

Query Match 100.0%; Score 41; DB 1; Length 686;  
Best Local Similarity 100.0%; Pred. No. 2.9e-34;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 TPFGKWPVPVFGRLASPGFGEYANDQERRWTLTAPGGR 41  
Db 16 TPFGKWPVPVFGRLASPGFGEYANDQERRWTLTAPGGR 56

Completed: January 11, 2003, 11:04:59  
e : 10 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 11:02:39 ; Search time 20 Seconds  
(without alignments)  
197.076 Million cell updates/sec

Title: US-09-874-198-1  
Perfect score: 41  
Sequence: 1 TPLGPKWPEPVFGRLASPGF.....GEYANDQERRWTLTAPPGYR 41

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
hed: 283224 seqs, 96134422 residues  
Size: 16

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : PIR\_73:\*  
1: plr1:\*  
2: plr2:\*  
3: plr3:\*  
4: plr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	686	1 A59271	Ra-reactive factor

#### ALIGNMENTS

#### RESULT 1

A59271  
Ra-reactive factor (EC 3.4.21.-) 2 precursor - human  
N:Alternate names: mannose binding protein-associated serine proteinase 2 (MASP-2)  
C:Species: Homo sapiens (man)  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 16-Jun-2000  
C:Accession: A59271  
R:Thiel, S.; Vorup-Jensen, T.; Stover, C.M.; Schwaible, W.J.; Laursen, S.B.; Poulsen, K.  
Nature 386, 506-510, 1997  
A:Title: A second serine protease associated with mannan-binding lectin that activates C  
A:Reference number: A59271; MUID:97242412; PMID:9087411  
A:Accession: A59271  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-686 <JEN>  
A:Cross-references: GB:Y09926; NID:g4007626; PIDN:CAW1059.1; PID:g4007627  
A:Experimental source: tissue liver  
A:Note: submitted to GenBank, December 1996  
A:Note: parts of this sequence, including the amino end of the mature protein, were dete  
C:Genetics:  
A:Gene: GDB:MASP2  
A:Cross-references: GDB:6071500  
A:Map position: 1p36.2-1p36.3

C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto  
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; hydrolase; serin  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-444,445-686/Product: Ra-reactive factor 2 #status predicted <MAT>  
F:19-134/Domain: C1r/C1s repeat homology <C1R1>  
F:142-180/Domain: EGF homology <EGF>  
F:184-293/Domain: C1r/C1s repeat homology <C1R2>  
F:300-361/Domain: complement factor H repeat homology <FH1>  
F:366-430/Domain: complement factor H repeat homology <FH2>  
F:445-679/Domain: trypsin homology <TRY>  
F:72-90,142-156,152-165,167-180,184-211,241-259,300-348,328-361,366-412,366-430,434-5  
F:158/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F:444-445/Cleavage site: Arg-Ile (autolytic) #status predicted  
F:483,532,633/Active site: His, Asp, Ser #status predicted

Query Match 100.0% Score 41: DB 1: Length 686;  
Best Local Similarity 100.0% Pred. No. 6.7e-35;  
Matches 41: Conservative 0: Mismatches 0: Indels 0: Gaps 0;  
OY 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGYR 41  
Db 16 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGYR 56

Search completed: January 11, 2003, 11:06:16  
Job time : 20 secs

**THIS PAGE BLANK (USPTO)**



```
XX Human mature MASP-2 protein.
DE
XX
XX Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KW Complement fixation; infection; microbe; retrovirus; HIV; abortion;
KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Region 15..671
FT /note="This region is specifically claimed in
FT claim 45"
FT
FT
PN WO200206460-A2.
XX
XX 24-JAN-2002.
XX
XX 13-JUL-2001; 2001WO-DK00499.
XX
XX 13-JUL-2000; 2000DK-0001089.
XX 01-JUN-2001; 2001DK-0000870.
XX
XX (JENS/) JENSENIUS J C.
XX (THIE/) THIEL S.
XX
XX Jensenius JC, Thiel S;
PI
XX WPI: 2002-179791/23.
XX
XX Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT infections
PT
XX Claim 45; Page 73-76; 76pp; English.
XX
XX The invention relates to use of a polypeptide derived from
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
CC producing a pharmaceutical composition. MASP-2 is a complement-
CC fixing enzyme and involved in lectin pathway of complement activation.
CC The pharmaceutical composition comprising MASP-2 is useful for
CC treating infections caused by microbes such as fungus, yeast,
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
CC bacteria which are resistant to at least one antibiotic medicament or
CC multiresistant. The polynucleotide encoding MASP-2 is useful for
CC treating patients deficient in MASP-2. The invention also discloses
CC MASP-2 assays which are useful for determination of MASP-2 activity or
CC levels in patients suffering from e.g. infections, inflammatory disorders
CC and spontaneous recurrent abortion. The pharmaceutical composition
CC comprising MASP-2 inhibitor is useful for treating inflammatory
CC disorders. The present sequence is human mature MASP-2 protein.
CC
XX
XX Sequence 671 AA;
SQ
XX
XX Query Match 100.0%; Score 41; DB 23; Length 671;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-33;
XX Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TPLGPKWPEVFGRLASPGFPGSEYANDQERRWTLTAPGGR 41
DB 1 TPLGPKWPEVFGRLASPGFPGSEYANDQERRWTLTAPGGR 41
XX
XX RESULT 3
XX ID AAE14564 standard; Protein: 686 AA.
XX AC AAE14564;
XX DT 17-MAY-2002 (first entry)
XX DE Human MASP-2 protein.
```

```
XX Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KW Complement fixation; infection; microbe; retrovirus; HIV; abortion;
KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..15
FT /label= Signal_peptide
FT Protein 16..686
FT /label= Mature_MASP-2_protein
FT Region 16..56
FT /note= "MASP-2 20 kDa fragment"
FT Domain 16..31
FT /label= C1r/C1s-like_domain
FT Region 16..296
FT /note= "This region is specifically claimed in claim 46"
FT Region 30..444
FT /note= "This region is specifically claimed in claim 41.
FT The region 30 to 296 is specifically claimed in claim 47"
FT Region 108..134
FT /note= "Fragment obtained by amino acid sequencing
FT of peptides"
FT Domain 135..146
FT /label= EGF-like_domain
FT Region 138..296
FT /note= "This region is specifically claimed in claim 43"
FT Misc-difference 155
FT /note= "Encoded by CA"
FT Misc-difference 156
FT /note= "Encoded by C"
FT Domain 183..199
FT /label= C1r/C1s-like_domain
FT Domain 293..307
FT /label= CCP-1_domain
FT /note= "Complement control protein domain"
FT Domain 363..376
FT /label= CCP-2_domain
FT Region 377..388
FT /note= "Fragment obtained by amino acid sequencing
FT of peptides"
FT Region 410..417
FT /note= "Fragment obtained by amino acid sequencing
FT of peptides"
FT Region 432..441
FT /note= "Linker"
FT Domain 445..468
FT /label= Serine_protease_domain
FT Active-site 483
FT Active-site 532
FT Active-site 633
XX
XX WO200206460-A2.
XX
XX 24-JAN-2002.
XX
XX 13-JUL-2001; 2001WO-DK00499.
XX
XX 13-JUL-2000; 2000DK-0001089.
XX 01-JUN-2001; 2001DK-0000870.
XX
XX (JENS/) JENSENIUS J C.
XX (THIE/) THIEL S.
XX
XX Jensenius JC, Thiel S;
XX WPI: 2002-179791/23.
XX N-PSDB: AAD24224.
XX
XX Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT
```



PT Infections -  
XX  
PS Claim 41; Fig 6; 76pp; English.  
CC  
XX The invention relates to use of a polypeptide derived from  
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for  
CC producing a pharmaceutical composition. MASP-2 is a complement-  
CC fixing enzyme and involved in lectin pathway of complement activation.  
CC The pharmaceutical composition comprising MASP-2 is useful for  
CC treating infections caused by microbes such as fungus, yeast,  
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic  
CC bacteria which are resistant to at least one antibiotic medicament or  
CC multiresistant. The polynucleotide encoding MASP-2 is useful for  
CC treating patients deficient in MASP-2. The invention also discloses  
CC MASP-2 assays which are useful for determination of MASP-2 activity or  
CC levels in patients suffering from e.g. infections, inflammatory disorders  
CC and spontaneous recurrent abortion. The pharmaceutical composition  
CC comprising MASP-2 inhibitor is useful for treating inflammatory  
CC disorders. The present sequence is human MASP-2 protein.  
CC Note: The present sequence is stated as being the same as  
CC SEQ ID NO:2 shown in sequence listing of the specification (AAE14568).  
XX However the sequences differ at various locations.  
XX  
SQ Sequence 686 AA:  
Query Match 100.0%; Score 41; DB 23; Length 686;  
Best Local Similarity 100.0%; Pred. No. 1.2e-33;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TPLGPKWPEPVFGRLASPCGFGEYANDQERRWTLTAPPGYR 41  
16 TPLGPKWPEPVFGRLASPCGFGEYANDQERRWTLTAPPGYR 56  
DB  
RESULT 4  
AAE14568  
ID AAE14568 standard; Protein; 686 AA.  
XX  
AC AAE14568;  
XX  
DT 17-MAY-2002 (first entry)  
XX  
DE Human MASP-2 protein, alternative version.  
XX  
KW Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;  
KW complement fixation; infection; microbe; retrovirus; HIV; abortion;  
KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.  
XX  
XX homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..15  
FT Protein 16..686  
FT Misc-difference 155 /label= Mature\_MASP-2\_protein  
FT Misc-difference 156 /note= "Encoded by CA"  
FT Misc-difference 156 /note= "Encoded by C"  
FT Misc-difference 298 /note= "Encoded by CAG"  
FT Misc-difference 299 /note= "Encoded by CCA"  
FT Misc-difference 299 /note= "Encoded by CCA"  
XX  
XX WO200206460-A2.  
XX  
XX 24-JAN-2002.  
XX  
XX 13-JUL-2001; 2001WO-DK00499.  
XX  
XX 13-JUL-2000; 2000DK-0001089.  
XX  
XX 01-JUN-2001; 2001DK-0000870.  
XX

PA (JENS/) JENSENIUS J C.  
PA (THIE/) THIEL S.  
XX  
XX  
PI Jensenius JC, Thiel S;  
XX  
XX WPI: 2002-179791/23.  
XX DR N-PSDB; AAD24224.  
XX  
PT Use of a polypeptide comprising amino acid sequence derived from  
PT mannan-binding lectin associated serine protease-2 (MASP-2) for  
PT producing pharmaceutical composition, to treat bacterial, fungal, viral  
PT infections -  
XX  
XX  
PS Claim 41; Page 71-73; 76pp; English.  
XX  
CC The invention relates to use of a polypeptide derived from  
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for  
CC producing a pharmaceutical composition. MASP-2 is a complement-  
CC fixing enzyme and involved in lectin pathway of complement activation.  
CC The pharmaceutical composition comprising MASP-2 is useful for  
CC treating infections caused by microbes such as fungus, yeast,  
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic  
CC bacteria which are resistant to at least one antibiotic medicament or  
CC multiresistant. The polynucleotide encoding MASP-2 is useful for  
CC treating patients deficient in MASP-2. The invention also discloses  
CC MASP-2 assays which are useful for determination of MASP-2 activity or  
CC levels in patients suffering from e.g. infections, inflammatory disorders  
CC and spontaneous recurrent abortion. The pharmaceutical composition  
CC comprising MASP-2 inhibitor is useful for treating inflammatory  
CC disorders. The present sequence is human MASP-2 protein.  
CC Note: The present sequence is stated as being the same as  
CC SEQ ID NO:2 shown in figure 6 of the specification (AAE14564).  
XX However the sequences differ at various locations.  
XX  
SQ Sequence 686 AA:  
Query Match 100.0%; Score 41; DB 23; Length 686;  
Best Local Similarity 100.0%; Pred. No. 1.2e-33;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TPLGPKWPEPVFGRLASPCGFGEYANDQERRWTLTAPPGYR 41  
16 TPLGPKWPEPVFGRLASPCGFGEYANDQERRWTLTAPPGYR 56  
DB  
RESULT 5  
ABG21134  
ID ABG21134 standard; Protein; 1034 AA.  
XX  
AC ABG21134;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #21125.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX

DR WPI: 2001-639362/73.  
DR N-PSDB: AAS85321.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX Claim 20: SEQ ID No 51493; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX SQ Sequence 1034 AA;

Query Match 100.0%; Score 41; DB 22; Length 1034;  
Best Local Similarity 100.0%; Pred. No. 1.7e-33;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPLGPKMPEPFGRLASPGFGEYANDQERRWTLTAPGVR 41  
|||||  
Db 752 TPLGPKMPEPFGRLASPGFGEYANDQERRWTLTAPGVR 792

Search completed: January 11, 2003, 11:04:36  
Job time : 36 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:51:49 : Search time 59 Seconds  
(without alignments)  
13.482 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPLGPKWPEPVFGRLASPGF.....GEYANDQERRWTLTAPGCR 41

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : PublishedApplications-AA:\*

- 1: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	41	10	US-09-874-198-1
2	238	100.0	41	10	US-09-874-238-1
3	238	100.0	686	10	US-09-874-198-2
4	238	100.0	686	10	US-09-874-238-2
5	93.5	39.3	174	10	US-09-764-833-363
6	93	39.1	679	10	US-09-874-198-6
7	93	39.1	679	10	US-09-874-238-6
8	91.5	38.4	525	9	US-10-004-551-18
9	91.5	38.4	525	9	US-10-004-551-20
10	91.5	38.4	525	9	US-10-004-551-22
11	91.5	38.4	526	9	US-10-004-551-24
12	91.5	38.4	526	9	US-10-004-551-14
13	91.5	38.4	526	9	US-10-004-551-16
14	81.5	34.2	144	9	US-09-886-429-4
15	81.5	34.2	144	9	US-10-041-406-5
16	81	34.0	101	10	US-09-799-118-6
17	81	34.0	351	10	US-09-925-302-611
18	81	34.0	449	10	US-09-919-497-89
19	81	34.0	688	10	US-09-874-198-7

118974 No. 5 hits  
are Applications  
in adequate  
conting.

20	81	34.0	688	10	US-09-874-238-7	Sequence 7, Appl1
21	81	34.0	705	9	US-09-808-602-94	Sequence 94, Appl1
22	78	32.8	458	10	US-09-925-301-1282	Sequence 1282, Ap
23	78	32.8	556	10	US-09-729-674-92	Sequence 92, Appl
24	74.5	31.3	230	10	US-09-925-300-1416	Sequence 1416, Ap
25	74.5	31.3	760	10	US-09-925-301-1024	Sequence 1024, Ap
26	73.5	30.9	673	10	US-09-874-198-8	Sequence 8, Appl1
27	73.5	30.9	673	10	US-09-874-238-8	Sequence 8, Appl1
28	72	30.3	116	9	US-10-041-406-4	Sequence 4, Appl1
29	70	29.4	997	10	US-09-747-371-3	Sequence 3, Appl1
30	69	29.0	101	9	US-09-887-593-6	Sequence 6, Appl1
31	69	29.0	113	9	US-09-852-2094-22	Sequence 22, Appl
32	69	29.0	113	9	US-10-086-623-22	Sequence 22, Appl
33	69	29.0	392	10	US-09-898-570-39	Sequence 39, Appl
34	69	29.0	730	10	US-09-850-048A-2	Sequence 2, Appl
35	69	29.0	986	10	US-09-285-385C-19	Sequence 19, Appl
36	69	29.0	999	10	US-09-747-371-2	Sequence 2, Appl
37	68	28.6	102	9	US-09-887-593-7	Sequence 7, Appl1
38	68	28.6	986	10	US-09-850-048A-4	Sequence 4, Appl1
39	68	28.6	1013	10	US-09-942-366-5	Sequence 5, Appl1
40	68	28.6	1013	10	US-09-285-385C-20	Sequence 20, Appl
41	68	28.6	1013	10	US-09-285-385C-2	Sequence 2, Appl1
42	67	28.2	1012	10	US-09-285-385C-4	Sequence 4, Appl1
43	67	28.2	1013	10	US-09-942-366-3	Sequence 3, Appl1
44	65.5	27.5	415	9	US-09-905-291A-104	Sequence 104, App
45	65.5	27.5	415	9	US-09-902-853-104	Sequence 104, App
46	65.5	27.5	415	9	US-09-807-824-104	Sequence 104, App
47	65.5	27.5	415	9	US-09-907-841-104	Sequence 104, App
48	65.5	27.5	415	9	US-09-904-011-104	Sequence 104, App
49	65.5	27.5	415	10	US-09-909-320-104	Sequence 104, App
50	65.5	27.5	415	10	US-09-909-088B-104	Sequence 104, App
51	63.5	26.7	459	10	US-09-789-561-97	Sequence 97, Appl
52	63	26.5	113	9	US-09-852-209A-23	Sequence 23, Appl
53	63	26.5	113	9	US-10-086-623-23	Sequence 23, Appl
54	63	26.5	923	12	US-10-104-440-2	Sequence 2, Appl1
55	63	26.5	931	12	US-10-104-440-4	Sequence 4, Appl1
56	60	25.2	77	12	US-10-067-422-13	Sequence 13, Appl
57	60	25.2	132	9	US-09-808-602-98	Sequence 98, Appl
58	60	25.2	270	9	US-09-732-242-2	Sequence 2, Appl1
59	60	25.2	396	10	US-09-800-729-86	Sequence 86, Appl
60	60	25.2	452	9	US-09-808-602-21	Sequence 21, Appl

## ALIGNMENTS

RESULT 1  
US-09-874-198-1.  
: Sequence 1, Application US/09874198  
: Patent No. US20020082208A1  
: GENERAL INFORMATION:  
: APPLICANT: Jensenius, Jens Chr.  
: APPLICANT: Thiel, Steffen  
: TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND  
: TITLE OF INVENTION: USES FOR IT  
: FILE REFERENCE: 09011-002002  
: CURRENT APPLICATION NUMBER: US/09/874,198  
: PRIOR FILING DATE: 2001-06-04  
: PRIOR APPLICATION NUMBER: 09/054,218  
: PRIOR FILING DATE: 1998-04-02  
: PRIOR APPLICATION NUMBER: 60/042,678  
: PRIOR FILING DATE: 1997-04-03  
: NUMBER OF SEQ ID NOS: 8  
: SOFTWARE: FASTSEQ for Windows Version 4.0  
: SEQ ID NO 1  
: LENGTH: 41  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-874-198-1  
Query Match 100.0%; Score 238; DB 10; Length 41;  
Best Local Similarity 100.0%; Pred. No. 5.1e-22;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```
FILE REFERENCE: 09011-002002
CURRENT APPLICATION NUMBER: US/09/874,198
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/054,218
PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 679
TYPE: PRT
ORGANISM: Homo sapiens
US-09-874-198-6

Query Match
Best Local Similarity 45.2%; Score 93; DB 10; Length 679;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

11 VGRGLASPGFPGGEYANDQERRWTLTAPPGYR 41
      |||: |||: | | | | | | | | | |
8 MFGQIQSPGYPGYPDSSEVTWNITVDPGFR 38

RESULT 7
US-09-874-238-6
Sequence 6, Application US/09874238
Patent No. US20020082209A1
GENERAL INFORMATION:
APPLICANT: Jensenius, Jens Chr.
APPLICANT: Thiel, Steffen
TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
FILE REFERENCE: 09011-002003
CURRENT APPLICATION NUMBER: US/09/874,238
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/054,218
PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 679
TYPE: PRT
ORGANISM: Homo sapiens
US-09-874-238-6

Query Match
Best Local Similarity 39.1%; Score 93; DB 10; Length 679;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

11 VGRGLASPGFPGGEYANDQERRWTLTAPPGYR 41
      |||: |||: | | | | | | | | | |
8 MFGQIQSPGYPGYPDSSEVTWNITVDPGFR 38

RESULT 8
US-10-004-551-18
Sequence 18, Application US/10004551
Publication No. US20030004310A1
GENERAL INFORMATION:
APPLICANT: SHINKETS, RICHARD A
APPLICANT: FERNANDES, ELMA
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 15966-559
CURRENT APPLICATION NUMBER: US/10/004,551
CURRENT FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 09/635,949
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 110
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
```

```
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
OTHER INFORMATION: n 2077 can be A, G, C, or T.
US-10-004-551-18

Query Match
Best Local Similarity 38.4%; Score 91.5; DB 9; Length 525;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

5 PKW--PEPV-----FGRLASPGFPGGEYANDQERRWTLTAPPGYR 41
      | | | | | | | | | | | | | | | | | | | | | |
403 PFWDKSEPVCIACGCGVINRGTGRIVSPGPGNYSNNLTCHWLLLEAPGQR 454

RESULT 9
US-10-004-551-20
Sequence 20, Application US/10004551
Publication No. US20030004310A1
GENERAL INFORMATION:
APPLICANT: SHINKETS, RICHARD A
APPLICANT: FERNANDES, ELMA
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 15966-559
CURRENT APPLICATION NUMBER: US/10/004,551
CURRENT FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 09/635,949
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 110
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 20
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
OTHER INFORMATION: n 2077 can be A, G, C, or T.
US-10-004-551-20

Query Match
Best Local Similarity 38.4%; Score 91.5; DB 9; Length 525;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

5 PKW--PEPV-----FGRLASPGFPGGEYANDQERRWTLTAPPGYR 41
      | | | | | | | | | | | | | | | | | | | | | |
403 PFWDKSEPVCIACGCGVINRGTGRIVSPGPGNYSNNLTCHWLLLEAPGQR 454

RESULT 10
US-10-004-551-22
Sequence 22, Application US/10004551
Publication No. US20030004310A1
GENERAL INFORMATION:
APPLICANT: SHINKETS, RICHARD A
APPLICANT: FERNANDES, ELMA
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 15966-559
CURRENT APPLICATION NUMBER: US/10/004,551
CURRENT FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 09/635,949
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 110
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
US-10-004-551-22

Query Match
Best Local Similarity 38.4%; Score 91.5; DB 9; Length 525;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

5 PKW--PEPV-----FGRLASPGFPGGEYANDQERRWTLTAPPGYR 41
      | | | | | | | | | | | | | | | | | | | | | |
```





```

; CURRENT APPLICATION NUMBER: US/09/874,198
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-874-198-7

```

```

Query Match          34.0%; Score 81; DB 10; Length 688;
Best Local Similarity 44.1%; Pred. No. 0.037;
Matches 15; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

```

```

QY 8 PEVYGRGLASPGPEGEYANDQERRMTLTAPPGYR 41
   1 : || : || || || : || : || || ||
   5 POKLFGVTSPLPFKPYPNNEFTTIVTPGGR 38

```

```

RESULT 20
US-09-874-238-7
; Sequence 7, Application US/09874238
; Patent No. US20020082209A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; TITLE OF INVENTION: USES FOR IT
; FILE REFERENCE: 09011-002003
; CURRENT APPLICATION NUMBER: US/09/874,238
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-874-238-7

```

```

Query Match          34.0%; Score 81; DB 10; Length 688;
Best Local Similarity 44.1%; Pred. No. 0.037;
Matches 15; Conservative 5; Mismatches 14; Indels 0; Gaps 0;
QY 8 PEVYGRGLASPGPEGEYANDQERRMTLTAPPGYR 41
   1 : || : || || || || : || : || || ||
   5 POKLFGVTSPLPFKPYPNNEFTTIVTPGGR 38

```

Search completed: January 11, 2003, 11:03:49  
Job time : 61 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:46:19 ; Search time 66 Seconds  
(without alignments)  
18.278 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238  
Sequence: 1 TPLGKPEPVGRLASPGF.....GEVANDERRMTLAPGVR 41

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/pdata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/pdata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/pdata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/pdata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/pdata/1/1aa/6CTUS.COMB.pep:\*  
6: /cgn2\_6/pdata/1/1aa/Backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83.5	35.1	1290	1	US-08-470-350B-2
2	81	34.0	101	2	US-08-242-097-6
3	81	34.0	101	2	US-09-206-695-6
4	81	34.0	449	2	US-08-839-008-9
5	78	32.8	449	2	US-08-839-008-2
6	76	31.9	52	4	US-09-227-357-363
7	74	31.1	401	2	US-08-839-008-5
8	74	31.1	468	2	US-08-839-008-7
9	74	31.1	468	4	US-09-032-523-8
10	69	29.0	101	4	US-09-374-135-6
11	69	29.0	730	4	US-08-872-757-2
12	69	29.0	788	1	US-08-572-225-1
13	68	28.6	102	4	US-09-374-135-7
14	68	28.6	591	3	US-08-991-408-4
15	68	28.6	591	3	US-09-432-473-4
16	68	28.6	986	4	US-08-872-757-4
17	68	28.6	1013	2	US-08-866-650-5
18	68	28.6	1013	2	US-08-866-650-5
19	68	28.6	1013	3	US-09-021-287-5
20	68	28.6	1013	3	US-08-991-408-2
21	68	28.6	1013	4	US-09-240-473-5
22	67	28.2	1013	4	US-09-432-473-2
23	67	28.2	1013	4	US-08-866-650-3
24	67	28.2	1013	4	US-09-021-287-3
25	65.5	27.5	415	4	US-09-240-473-3
26	64	26.9	925	4	US-09-032-523-2
27	63	26.5	532	1	US-08-494-168-9

*Not a sig w gland*

28	63	26.5	909	3	US-08-936-135-18	Sequence 18, App1
29	63	26.5	922	4	US-09-116-473-4	Sequence 4, App1
30	63	26.5	923	3	US-08-936-135-6	Sequence 6, App1
31	63	26.5	926	3	US-08-936-135-20	Sequence 20, App1
32	61	25.6	901	3	US-08-936-135-22	Sequence 22, App1
33	61	25.6	906	3	US-08-936-135-24	Sequence 24, App1
34	61	25.6	909	3	US-08-936-135-8	Sequence 8, App1
35	61	25.6	909	3	US-08-936-135-10	Sequence 10, App1
36	61	25.6	914	3	US-08-936-135-12	Sequence 12, App1
37	61	25.6	926	3	US-08-936-135-14	Sequence 14, App1
38	61	25.6	931	3	US-08-936-135-16	Sequence 16, App1
39	60	25.2	1057	3	US-08-931-820-4	Sequence 4, App1
40	60	25.2	1078	3	US-08-963-825-21	Sequence 21, App1
41	60	25.2	1078	4	US-09-500-811-21	Sequence 21, App1
42	60	25.2	1078	4	US-09-570-573-21	Sequence 21, App1
43	60	25.2	1078	4	US-09-548-608-21	Sequence 21, App1
44	58	24.4	666	4	US-09-341-587-1	Sequence 1, App1
45	58	24.4	1785	4	US-09-341-587-3	Sequence 1, App1
46	57.5	24.2	277	2	US-08-024-868-2	Sequence 2, App1
47	57.5	24.2	277	2	US-08-242-097-2	Sequence 2, App1
48	57.5	24.2	277	4	US-09-206-695-2	Sequence 2, App1
49	57.5	24.2	277	4	US-09-000-179-1	Sequence 1, App1
50	57.5	24.2	277	5	PCT-US96-11995-1	Sequence 1, App1
51	57.5	24.2	546	1	US-08-494-168-10	Sequence 10, App1
52	54.5	22.9	446	2	US-08-922-171-3	Sequence 3, App1
53	54.5	22.9	472	2	US-08-922-171-2	Sequence 2, App1
54	54	22.7	39	1	US-08-534-342-5	Sequence 5, App1
55	54	22.7	39	1	US-08-534-342-6	Sequence 6, App1
56	54	22.7	39	1	US-08-675-140-5	Sequence 5, App1
57	54	22.7	39	1	US-08-675-140-6	Sequence 6, App1
58	53	22.3	250	4	US-09-134-001C-2869	Sequence 2869, App1
59	53	22.3	351	1	US-08-470-202-46	Sequence 46, App1
60	53	22.3	351	1	US-08-471-770-46	Sequence 46, App1

## ALIGNMENTS

RESULT 1  
US-08-470-350B-2  
; Sequence 2, Application US/08470350B  
; Patent No. 5684126  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xiao  
; ATTORNEY: Snyder, Solomon H  
; TITLE OF INVENTION: Ebnerin: A Secreted von Ebner's Gland  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08470,350B  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wolfe, Susan A  
; REGISTRATION NUMBER: 33,568  
; REFERENCE/DOCKET NUMBER: 01107.48790  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 1290 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-470-350B-2

Query Match  
Best Local Similarity 35.1%; Score 83.5; DB 1; Length 1290;  
Matches 16; Conservative 5; Mismatches 17; Indels 13; Gaps 1;

QY 4 GPKWPEPVF-----GRLASPGFPGGEYANDQERRMTLTAPPGYR 41  
Db 825 GPWMTSPFVNYTCGGFLTGLSQGFSSPYPGSYPNNAACLMNIEVPPNRYR 875

RESULT 2  
US-08-242-097-6  
Sequence 6, Application US/08242097

PATENT No. 5846763  
GENERAL INFORMATION:  
APPLICANT: Lee, Tae Ho  
ATTORNEY/AGENT INFORMATION:  
NAME: Wisniewski, Hans Georg  
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding  
TITLE OF INVENTION: Therefore and Uses Thereof  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/242,097  
FILING DATE: 13-MAY-1994  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-628-5197  
TELEFAX: 212-737-3528

ORAMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 101 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-242-097-6

Query Match  
Best Local Similarity 34.0%; Score 81; DB 2; Length 101;  
Matches 15; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 8 PEPVGRGLASPGFPGGEYANDQERRMTLTAPPGYR 41  
Db 7 POKLFGFVTSPLFKPKYPNNFETTVITVPTGYR 40

RESULT 3  
US-09-206-695-6  
Sequence 6, Application US/09206695  
PATENT No. 6210905  
GENERAL INFORMATION:  
APPLICANT: Lee, Tae Ho  
ATTORNEY/AGENT INFORMATION:  
NAME: Wisniewski, Hans Georg

APPLICANT: Vilcek, Jan  
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding  
TITLE OF INVENTION: Therefore and Uses Thereof  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/206,695  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/242,097  
FILING DATE: 13-MAY-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-628-5197  
TELEFAX: 212-737-3528

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 101 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-206-695-6

Query Match  
Best Local Similarity 34.0%; Score 81; DB 4; Length 101;  
Matches 15; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 8 PEPVGRGLASPGFPGGEYANDQERRMTLTAPPGYR 41  
Db 7 POKLFGFVTSPLFKPKYPNNFETTVITVPTGYR 40

RESULT 4  
US-08-839-008-9  
Sequence 9, Application US/08839008

PATENT No. 5916758  
GENERAL INFORMATION:  
APPLICANT: Hurtle, Mark R.  
APPLICANT: McDonnell, Peter C.  
APPLICANT: McNulty, Dean E.  
APPLICANT: Rosen, Craig A.  
APPLICANT: Siemens, Ivo R.  
APPLICANT: Young, Peter R.  
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithkline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/839,008  
;; FILING DATE: 23-APR-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/563,697  
;; FILING DATE: 28-NOV-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Baumeister, Kirk  
;; REGISTRATION NUMBER: 33,833  
;; REFERENCE/DOCKET NUMBER: P50384  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 610-270-5096  
;; TELEFAX: 610-270-5090  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 449 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-839-008-9  
;  
Query Match 34.0%; Score 81; DB 2; Length 449;  
Best Local Similarity 37.8%; Pred. No. 0.014;  
Matches 17; Conservative 5; Mismatches 13; Indels 10; Gaps 1;  
;  
QY 5 PKWPEPVF-----GRLASPGFGEXANDQERRMTLTAPG 39  
1: 111 1:11111 1:111111  
DB 28 PNTRPVFLCGGDVKGSGYVASEGFPNSYPPNKECIWITVPEG 72

RESULT 5  
US-08-839-008-2  
Sequence 2, Application US/08839008  
Patent No. 5916758  
GENERAL INFORMATION:  
APPLICANT: Hurler, Mark R  
APPLICANT: McDonnell, Peter C  
APPLICANT: McNulty, Dean E  
APPLICANT: Rosen, Craig A  
APPLICANT: Siemens, Ivo R  
APPLICANT: Young, Peter R  
APPLICANT: Yue, Tian-li  
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithkline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/839,008  
FILING DATE: 23-APR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/563,697  
FILING DATE: 28-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: P50384  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
INFORMATION FOR SEQ ID NO: 2:  
;

;;  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 449 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-839-008-2  
;  
Query Match 32.8%; Score 78; DB 2; Length 449;  
Best Local Similarity 37.8%; Pred. No. 0.034;  
Matches 17; Conservative 5; Mismatches 13; Indels 10; Gaps 1;  
;  
QY 5 PKWPEPVF-----GRLASPGFGEXANDQERRMTLTAPG 39  
1: 111 1:11111 1:111111  
DB 28 PNTRPVFLCGGDVKGSGYVASEGFPNLYPPNKECIWITVPEG 72  
;  
RESULT 6  
US-09-227-357-363  
Sequence 363, Application US/09227357  
Patent No. 6342381  
GENERAL INFORMATION:  
APPLICANT: Fischer et al.  
TITLE OF INVENTION: 123 Human Secreted Proteins  
FILE REFERENCE: P2010P1  
CURRENT APPLICATION NUMBER: US/09/227,357  
EARLIER FILING DATE: 1999-01-08  
EARLIER APPLICATION NUMBER: PCT/US98/13684  
EARLIER FILING DATE: 1998-07-07  
EARLIER APPLICATION NUMBER: 60/051,926  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,793  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,925  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,929  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,803  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,732  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,931  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,932  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,916  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,930  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,918  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
;

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-Apr-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-Nov-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SPO ID NO: 7:
LENGTH: 468 amino acids

```

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-839-008-7

Query Match 31.1%; Score 74; DB 2; Length 468;  
Best Local Similarity 35.6%; Pred. No. 0.11;  
Matches 16; Conservative 6; Mismatches 13; Indels 10; Gaps 1;

QY 5 PKMPEPVF-----GRLASPGFPGGEYANDQERRMTLTAPPG 39  
1 : 111 1 : 111 11 : 11:1 1  
Db 27 PNTRPVFLCGDVTGSGYVASEGPNLYPPNKKCIWITVPEG 71

RESULT 9  
US-09-032-523-8  
Sequence 8, Application US/09032523  
Patent No. 6232454

GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl  
APPLICANT: Baugh, Mariah  
TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,523  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0479 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 468 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 2589009  
US-09-032-523-8

Query Match 31.1%; Score 74; DB 4; Length 468;  
Best Local Similarity 35.6%; Pred. No. 0.11;  
Matches 16; Conservative 6; Mismatches 13; Indels 10; Gaps 1;

QY 5 PKMPEPVF-----GRLASPGFPGGEYANDQERRMTLTAPPG 39  
1 : 111 1 : 111 11 : 11:1 1  
Db 27 PNTRPVFLCGDVTGSGYVASEGPNLYPPNKKCIWITVPEG 71

RESULT 10  
US-09-374-135-6  
Sequence 6, Application US/09374135  
Patent No. 6277972

GENERAL INFORMATION:  
APPLICANT: Afar, Daniel E.  
APPLICANT: Hubert, Rene S.  
APPLICANT: Leong, Kahan  
APPLICANT: Raitano, Arthur B.  
APPLICANT: Safiran, Douglas C.  
APPLICANT: Jakobovits, Aya  
TITLE OF INVENTION: BPC-1: A SECRETED BRAIN-SPECIFIC PROTEIN EXPRESSED AND  
FILE REFERENCE: 1703-017 US1  
TITLE OF INVENTION: SECRETED BY PROSTATE AND BLADDER CANCER CELLS  
CURRENT APPLICATION NUMBER: US/09/374,135  
CURRENT FILING DATE: 1999-08-10  
PRIOR APPLICATION NUMBER: 60/095,982  
PRIOR FILING DATE: 1998-08-10  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 101  
TYPE: PRT  
ORGANISM: Mouse

Query Match 29.0%; Score 69; DB 4; Length 101;  
Best Local Similarity 44.8%; Pred. No. 0.089;  
Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 13 GRLASPGFPGGEYANDQERRMTLTAPPGR 41  
1 : 111:1 11 : 11 11 11  
Db 1 GSTSPGMPKPEYPPNKNCIWOLVAPTOYR 29

RESULT 11  
US-08-872-757-2  
Sequence 2, Application US/08872757  
Patent No. 6238584

GENERAL INFORMATION:  
APPLICANT: Prockop, Darwin J.  
APPLICANT: Hojima, Yoshio  
APPLICANT: Li, Shi-Wu  
APPLICANT: Sieron, Aleksander  
TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/872,757  
FILING DATE: 10-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/609,187  
FILING DATE: 01-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P.  
REGISTRATION NUMBER: 25,227  
REFERENCE/DOCKET NUMBER: 8389-028-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-854-3660

```

:
: TELEFAX: 415-854-1694
:
: TELEX: 66141 PENNIE
:
: INFORMATION FOR SEQ ID NO: 2
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 730 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
:
: US-08-872-757-2

```

Query Match 29.0%; Score 69; DB 4; Length 730;  
Best Local Similarity 44.8%; Pred. No. 0.8;  
Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 13 GRLSPGPGGEYANDQERRWTLTAPPGYR 41  
| : ||| : | : | | | | |  
Db 600 GSITSPGWPKEYPPNKNCIQLVAPTQYR 628

T 12  
572-225-1  
idence 1, Application US/08572225  
tent No. 5807981

1 GENERAL INFORMATION:  
2 - APPLICANT: Prockop, Darwin J.  
3 APPLICANT: Hojima, Yoshio  
4 APPLICANT: Li, Shi-wu  
5 APPLICANT: Sieron, Aleksander  
6 APPLICANT: Brenner, Mitch  
7 TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND ITS USE FOR  
8 TITLE OF INVENTION: DRUG DEVELOPMENT FOR THE TREATMENT OF DISEASE  
9 - NUMBER OF SEQUENCES: 24  
10 CORRESPONDENCE ADDRESS:

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patentin Release #1.0, Version #1.30
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/572,225
8  FILING DATE: 13-DEC-1995

```

ATTORNEY/AGENT INFORMATION:  
NAME: Hallin, Albert P.  
REGISTRATION NUMBER: 25,227  
REFERENCE/DOCKET NUMBER: 8389-031  
TELECOMMUNICATION INFORMATION:

SEQUENCE CHARACTERISTICS:  
LENGTH: 788 amino acids  
TYPE: amino acid  
STRADEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-572-225-1

Query Match 29.0%; Score 69; DB 1; Length 788;  
Best Local Similarity 44.8%; Pred. NO. 0.88;  
Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

```

QY      13  GRLASPGFPGEYANDQERRWTLTAPPGR 41
          | : ||| : | : | | | |
Db      402  GSTSPGWPKEYPPNKNCIWLVAFTQYR 430

```

RESULT 13  
US-09-374-135-7  
; Sequence 7, Application US/09374135

GENERAL INFORMATION:  
 APPLICANT: Afar, Daniel E.  
 APPLICANT: Hubert, Rene S.  
 APPLICANT: Leong, Kahan  
 APPLICANT: Raitano, Arthur B.  
 APPLICANT: Saffran, Douglas C.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

```

; SEQ ID NO 1
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Mouse
US-09-374-135-7

```

query Match	28.6%;	Score 68;	DB 4;	Length 102;
Best Local Similarity	34.5%;	Pred. No. 0.12;		
Matches	10;	Conservative	9;	Mismatches 10;
			Indels	0;
			Gaps	0;

```
QY      13  GRLSPGPFGEYANDQERRWLTAPPGR  41
          | : | | : | : | : | : | : |
Db      1  GTITSPNWPDKYPSKKECTWAISSTPGHR  29
```

RESULT 14  
US-08-991-408-4  
; Sequence 4, Application US/08991408  
; Patent No. 6008017

GENERAL INFORMATION:  
APPLICANT: ARLETH, ANTHONY J.  
APPLICANT: WILLEITE, ROBERT N.  
APPLICANT: ELSHOURBAGY, NABIL A.  
APPLICANT: LI, XIAORONG  
TITLE OF INVENTION: HUAN CARDIAC/BRAIN TOLLID-LIKE  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 4

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Diskette
3  COMPUTER: IBM Compatible
4  OPERATING SYSTEM: DOS
5  SOFTWARE: FASTSEQ FOR Windows Version 2.0
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/991,408

```

CLASSIFICATION: PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/034,471  
FILING DATE: 02-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL E.  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: ATG-50038  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700

TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 591 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-991-408-4

Query Match 28.6%; Score 68; DB 3; Length 591;  
Best Local Similarity 37.9%; Pred. No. 0.85;  
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 13 GRLASPGFGEYANDQERRWLTAPPQYR 41  
DB 361 GLITSPNMPDKTPSRKECTWEISATPGHR 389

15  
432-473-4  
Sequence 4, Application US/09432473  
Patent No. 6365715  
GENERAL INFORMATION:  
APPLICANT: ARLETH, ANTHONY J.  
APPLICANT: WILLETT, ROBERT N.  
APPLICANT: ELSHOURBAGY, MABIL A.  
APPLICANT: LI, XIATONG  
TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLOID-LIKE PROTEIN  
FILE REFERENCE: ATG-50038-D1  
CURRENT APPLICATION NUMBER: US/09/432,473  
EARLIER FILING DATE: 1999-11-01  
EARLIER APPLICATION NUMBER: 08/991,408  
EARLIER FILING DATE: 1997-12-16  
EARLIER APPLICATION NUMBER: 60/034,471  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 591  
TYPE: PRT  
ORGANISM: HOMO SAPIENS  
US-09-432-473-4

Query Match 28.6%; Score 68; DB 4; Length 591;  
Best Local Similarity 37.9%; Pred. No. 0.85;  
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

13 GRLASPGFGEYANDQERRWLTAPPQYR 41  
DB 361 GLITSPNMPDKTPSRKECTWEISATPGHR 389

RESULT 16  
US-08-872-757-4  
Sequence 4, Application US/08872757  
Patent No. 6258584  
GENERAL INFORMATION:  
APPLICANT: Prockop, Darwin J.  
APPLICANT: Hojima, Yoshio  
APPLICANT: LI, Shi-Wu  
APPLICANT: Sieron, Aleksander  
TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND  
NUMBER OF INVENTION: PROCESSES, METHODS AND USES THEREOF  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/872,757  
FILING DATE: 10-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/609,187  
FILING DATE: 01-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P.  
REGISTRATION NUMBER: 25,227  
REFERENCE/DOCKET NUMBER: 8389-028-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-854-3660  
TELEFAX: 415-854-3694  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 986 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-872-757-4

Query Match 28.6%; Score 68; DB 4; Length 986;  
Best Local Similarity 34.5%; Pred. No. 1.5;  
Matches 10; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 13 GRLASPGFGEYANDQERRWLTAPPQYR 41  
DB 756 GTITSPNMPDKTPSRKECTWEISATPGHR 784

RESULT 17  
US-08-866-650-5  
Sequence 5, Application US/08866650  
Patent No. 5939321  
GENERAL INFORMATION:  
APPLICANT: Greenspan, Daniel S  
APPLICANT: Takahara, Kazuhiko  
APPLICANT: Hoffman, Guy G  
TITLE OF INVENTION: Mammalian Tolloid-Like Protein  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: Quarles & Brady  
STREET: 1 South Plackney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/866,650  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Berson, Bennett J  
REGISTRATION NUMBER: 37094  
REFERENCE/DOCKET NUMBER: 960296, 93839  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1013 amino acids







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:44:28 : Search time 66 Seconds  
(without alignments)  
59.720 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPLGPKMPEPVFGRLASPGF.....GEYANDQERRWTLTAPGVR 41

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ed: 283224 seqs, 96134422 residues

number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	238	100.0	686	1 A59271	Ra-reactive factor
2	93	39.1	699	1 I54763	hypothetical factor
3	85	35.7	610	2 T16761	hypothetical prote
4	83.5	35.1	1290	2 A51190	ebnerin precursor
5	81	34.0	449	2 A55362	procollagen I C-pr
6	81	34.0	705	1 C1HURB	complement subcomp
7	80.5	33.8	977	2 I52657	seizure-related pr
8	78	32.8	3623	2 T09456	intrinsic factor-B
9	74.5	31.3	688	1 C1HUS	complement subcomp
10	74	31.1	319	2 I51569	UVS-2 protein - A1
11	74	31.1	402	2 JH0403	procollagen I C-pr
12	72.5	30.5	695	1 S05008	complement subcomp
13	72	30.3	3623	2 T08618	intrinsic factor-B
14	71.5	30.0	1070	2 T31069	tolloid-BMP-1 like
15	69	29.0	730	1 BKHU1	procollagen C-endo
16	69	29.0	823	1 A58788	procollagen C-endo
17	69	29.0	986	1 B58788	procollagen C-endo
18	69	29.0	991	2 I49540	procollagen C-endo
19	69	29.0	1464	2 S58984	development protei
20	68	28.6	504	2 S56745	mucin (clone PG31
21	68	28.6	767	2 T30018	hypothetical prote
22	68	28.6	1594	2 T30549	hensin - rabbit
23	68	28.6	2083	2 T42721	CRP-ductin-alpha p
24	67.5	28.4	694	2 T65554	complement subcomp
25	67	28.2	288	2 T33224	hypothetical prote
26	67	28.2	707	2 J03218	procollagen C-endo
27	65.5	27.5	77	2 S10114	hypothetical prote
28	65.5	27.5	171	2 A34493	collagen alpha 1(I
29	65.5	27.5	920	2 B34493	collagen alpha 1(I

*Hit No. 1 is  
Redundant  
Guess Protein  
SPTRRBLR  
results.*

30	65	27.3	1057	1 A39288	dorsal-ventral pat
31	65	27.3	3871	2 T22812	hypothetical prote
32	63	26.5	343	2 T20614	hypothetical prote
33	63	26.5	391	2 T34284	hypothetical prote
34	63	26.5	1524	2 T30337	polyprotein - Afri
35	63	26.5	1669	1 CGH04B	conserved hypochat
36	62	26.1	980	2 E72035	CT590 hypochat
37	62	26.1	980	2 E86589	hypothetical prote
38	61	25.6	294	2 T29838	hypothetical prote
39	60.5	25.4	271	2 I51536	XeLMPa.aa - Africa
40	60.5	25.4	886	2 I50694	collagen alpha 1(I
41	60	25.2	418	2 G97360	intracellular PHB
42	60	25.2	418	2 A82578	collagen alpha 2 c
43	60	25.2	1466	1 CGH07L	hypothetical prote
44	59.5	25.0	1433	2 A46053	collagen alpha 1(I
45	59	24.8	511	2 G95394	bullos pemphigoid
46	59	24.8	636	2 S41067	probable oxygenase
47	59	24.8	1051	2 A35763	collagen alpha 1(I
48	58.5	24.6	161	2 A83269	hypothetical prote
49	58	24.4	105	2 T18036	hypothetical prote
50	58	24.4	347	2 T20618	hypothetical prote
51	58	24.4	381	2 T20622	hypothetical prote
52	58	24.4	1464	2 S59856	collagen alpha 1(I
53	58	24.4	1670	1 CGH03B	collagen alpha 3(I
54	58	24.4	2403	2 A59386	sanko - human
55	57.5	24.2	223	2 T16654	hypothetical prote
56	57.5	24.2	275	2 J06506	tumor necrosis fac
57	57.5	24.2	277	2 A41735	hyaluronate-bindin
58	57.5	24.2	349	2 T29931	hypothetical prote
59	57.5	24.2	1497	2 I49607	procollagen type V
60	57.5	24.2	1691	1 S22917	collagen alpha 5(I

## ALIGNMENTS

RESULT 1  
A59271  
Ra-reactive factor (EC 3.4.21.-) 2 precursor - human  
N:Alternate names: mannose binding protein-associated serine proteinase 2 (MASP-2)  
C:Species: Homo sapiens (man)  
C:Date: 19-May-2000 #sequence-Revision 19-May-2000 #text-change 16-Jun-2000  
R:Thiel, S.; Vorup-Jensen, T.; Stover, C.M.; Schwaeble, W.J.; Laursen, S.B.; Poulsen, Nature 386, 506-510, 1997  
A>Title: A second serine protease associated with mannan-binding lectin that activate  
A:Reference number: A59271; M0ID:97242412; PMID:9087411  
A:Accession: A59271  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-686 <REN>  
A:Cross-references: GB:Y09926; NID:94007626; PIDN:CAAF1059.1; PID:94007627  
A:Experimental source: tissue liver  
A>Note: submitted to GenBank, December 1996  
A>Note: parts of this sequence, including the amino end of the mature protein, were d  
C:Genetics:  
A:Gene: GDB:MASP2  
A:Cross-references: GDB:6071500  
A:Map position: 1p36.2-1p36.3  
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto  
C:Keywords: beta-hydroxyparagline; complement pathway; duplication; hydrolase; serin  
F:1-15/Domain: signal sequence #status predicted <StC>  
F:16-444/Domain: 445-686/Product: Ra-reactive factor 2 #status predicted <MAT>  
F:19-134/Domain: C1r/C1s repeat homology <C1R1>  
F:142-180/Domain: EGF homology <EGF>  
F:184-293/Domain: C1r/C1s repeat homology <C1R2>  
F:360-361/Domain: complement factor H repeat homology <FH1>  
F:366-430/Domain: complement factor H repeat homology <FH2>  
F:445-679/Domain: trypsin homology <TRY>  
F:72-90/142-156/152-165/167-180/184-211/241-259/300-348/328-361/366-412/396-430/434-5  
F:158/Modified site: erythro-beta-hydroxyparagline (Asn) #status predicted  
F:444-445/Cleavage site: Arg-Ile (autolytic) #status predicted  
F:483/532/633/Active site: His, Asp, Ser #status predicted

Query Match 100.0%; Score 238; DB 1; Length 686;  
Best Local Similarity 100.0%; Pred. No. 4.7e-21;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPLGPKWPEPVFGRLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 16 TPLGPKWPEPVFGRLASPGFPGGYANDQERRMTLTAPPGR 56

## RESULT 2

Re-activative factor (EC 3.4.21.-) 1 precursor - human  
N:Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)  
C:Species: Homo sapiens (man)  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 16-Jun-2000

C:Accession: I54763; JN0883  
R:Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.  
Int. Immunol. 6, 665-669, 1994

File: Molecular characterization of a novel serine protease involved in activation of  
Reference number: I54763; MUID:94289349; PMID:8018603  
Accession: I54763

File: Preliminary: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-699 <SAR>  
A:Cross-references: GB:D28593; NID:9790963; PIDN:BA05928.1; PID:9471128

R:Takada, F.; Takayama, Y.; Hattuse, H.; Kawakami, M.  
Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993

A:Title: A new member of the C1s family of complement proteins found in a bactericidal  
A:Reference number: JN0883; MUID:94059062; PMID:8240317

A:Accession: JN0883

A:Molecule type: mRNA

A:Residues: 1-234, 'E', 236-284, 'G', 286-498, 'K', 500-542, 'K', 544-642, 'S', 644-699 <TAK>  
A:Cross-references: DBJ:017325; NID:9439712; PIDN:BA04477.1; PID:9439713

A:Experimental source: liver

C:Comment: This is a serum bactericidal factor that activates complement C4 and C2 compo

C:Genetics:

A:Gene: GDB:MASP1; GDB:CRARF; CRARF1; PRSS5; MASP  
A:Cross-references: GDB:361104; GDB:330954; OMIM:600521

A:Map position: 3q27-3q28

C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H  
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hydro

F:1-11/Domain: Signal sequence #status predicted <SIS>  
F:18-448/449-699/Product: Ra-reactive factor #status predicted <ANT>

F:19-135/Domain: C1r/C1s repeat homology <C1R1>  
F:143-181/Domain: EGF homology <EGF>

F:185-294/Domain: C1r/C1s repeat homology <C1R2>  
F:301-362/Domain: complement factor H repeat homology <RH1>

F:432/Domain: complement factor H repeat homology <RH2>  
F:691/Domain: trypsin homology <TRY>

F:178/407/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:159/91/143-157/153-166/168-181/185-212/242-260/301-349/329-362/367-414/397-432/436-572,

F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F:448-449/Cleavage site: Arg-ile (autolytic) #status predicted

F:490/552/646/Active site: His, Asp, Ser #status predicted

Query Match 39.1%; Score 93; DB 1; Length 699;  
Best Local Similarity 45.2%; Pred. No. 0.0019;

Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 11 VGRGLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 27 MFGQIDSPGYPDPSYSDSEVTWITVPDGR 57

QY 11 VGRGLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 27 MFGQIDSPGYPDPSYSDSEVTWITVPDGR 57

## RESULT 3

hypothetical protein R151.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 23-Mar-2001

C:Accession: T16761

R:Fulton, L.

submitted to the EMBL Data Library, February 1994

A:Description: The sequence of C. elegans cosmid R151.  
A:Reference number: S44639

A:Accession: T16761

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-610 <FUI>

A:Cross-references: EMBL:000036; NID:9458996; PID:9459001; PIDN:AAA50653.1; CESP:R151

C:Genetics:

A:Gene: CESP:R151.5

A:Introns: 68/3; 120/1; 151/1; 307/1; 448/2; 525/2; 567/2; 592/2

Query Match 35.7%; Score 85; DB 2; Length 610;  
Best Local Similarity 51.9%; Pred. No. 0.015;

Matches 14; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 14 RLASPGFPGGYANDQERRMTLTAPPGR 40  
DB 383 RLSPGYPREFEGQCSWLTWAPPGR 409

## RESULT 4

A57190  
ebnerin precursor - rat

C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 05-Nov-1999

C:Accession: A57190

R:Li, X.T.; Snyder, S.H.

J. Biol. Chem. 270, 17674-17679, 1995

A:Title: Molecular cloning of Ebnerin, a von Ebner's gland protein associated with ta

A:Reference number: A57190; MUID:9535352; PMID:7629065

A:Accession: A57190

A:Molecule type: mRNA

A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Residues: 1-1290 <LTA>

A:Cross-references: GB:U32681; NID:97975346; PIDN:AA552248.1; PID:97975347

C:Superfamily: scavenger receptor cysteine-rich domain homology; C1r/C1s repeat homol

C:Keywords: extracellular protein; transmembrane protein

F:1-77/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC

F:142-245/Domain: scavenger receptor cysteine-rich domain homology <SRC1>  
F:289-397/Domain: scavenger receptor cysteine-rich domain homology <SRC2>

F:583-688/Domain: C1r/C1s repeat homology <C1R1>  
F:713-816/Domain: scavenger receptor cysteine-rich domain homology <SRC3>

F:838-944/Domain: C1r/C1s repeat homology <C1R2>

Query Match 35.1%; Score 83.5; DB 2; Length 1290;  
Best Local Similarity 31.4%; Pred. No. 0.053; 17; Indels 13; Gaps 1;  
Matches 16; Conservative 5; Mismatches 13; Indels 13; Gaps 1;

QY 4 GPKWPEPV-----GRLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 825 GPMVNSPPVNTVTCGFLGLSGQSSPYPSYPMNACLNIEVPNNR 875

QY 4 GPKWPEPV-----GRLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 825 GPMVNSPPVNTVTCGFLGLSGQSSPYPSYPMNACLNIEVPNNR 875

QY 4 GPKWPEPV-----GRLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 825 GPMVNSPPVNTVTCGFLGLSGQSSPYPSYPMNACLNIEVPNNR 875

QY 4 GPKWPEPV-----GRLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 825 GPMVNSPPVNTVTCGFLGLSGQSSPYPSYPMNACLNIEVPNNR 875

QY 4 GPKWPEPV-----GRLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 825 GPMVNSPPVNTVTCGFLGLSGQSSPYPSYPMNACLNIEVPNNR 875

QY 4 GPKWPEPV-----GRLASPGFPGGYANDQERRMTLTAPPGR 41  
DB 825 GPMVNSPPVNTVTCGFLGLSGQSSPYPSYPMNACLNIEVPNNR 875

C:Superfamily:11r/C1s repeat homology  
 C:Keywords: extracellular protein; glycoprotein; pyroglutamic acid  
 F:1-25/Domain: signal sequence #status predicted <StG>  
 F:26-449/Product: #status predicted <MAT>  
 F:7-146/Domain: C1r/C1s repeat homology <C1R1>  
 F:159-270/Domain: C1r/C1s repeat homology <C1R2>  
 F:29/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:29,431/Binding site: carbohydurate (Asn) (covalent) #status predicted

Query Match 34.0%; Score 81; DB 2; Length 449;  
 Best Local Similarity 37.8%; Pred. No. 0.034;  
 Matches 17; Conservative 5; Mismatches 13; Indels 10; Gaps 1;

QY 5 PKMPEPVF-----GRASPGFGEYANDQERRMTLTTPPG 39  
 Db 28 PNTRPVFLCGDVGKSGYVASEGFPNSYPNKECITWTIVPEG 72  
 T 6  
 Element subcomponent C 1RBA.R.GIF (EC 3.4.21.41) precursor [validated] - human  
 Species: Homo sapiens (man)  
 C.Date: 15-Nov-1984 #sequence\_revision 30-Jun-1991 #text\_change 03-Jun-2002  
 C.Accession: A241170; A29768; A29769; S02422; A00916; A37820; S68830  
 R.Leytus, S.P.; Kurachi, K.; Sakariassen, K.S.; Davie, E.W.  
 Biochemistry 25, 4855-4863, 1986  
 A>Title: Nucleotide sequence of the cDNA coding for human complement C1r.  
 A.Reference number: A24170; PMID:87026566; PMID:3021205  
 A.Accession: A24170  
 A.Molecule type: mRNA  
 A.Residues: 1-705 <KEY>  
 A:Cross-references: GB:M14058; NID:g179643; PIDN:AAA51851.1; PID:g179644  
 R.Journeel, A.; Tosi, M.  
 Biochem. J. 240, 783-787, 1986  
 A>Title: Cloning and sequencing of full-length cDNA encoding the precursor of human comp  
 A.Reference number: A29768; PMID:87156625; PMID:3030286  
 A.Accession: A29768  
 A.Molecule type: mRNA  
 A.Residues: 1-151, 'L', 153-705 <JOU>  
 A:Cross-references: GB:X04701; NID:g29538; PIDN:CMA28407.1; PID:g29539  
 R.Arlaud, G.J.; Willis, A.C.; Gagnon, J.  
 Biochem. J. 241, 711-720, 1987  
 A>Title: Complete amino acid sequence of the A chain of human complement-classical-pathw  
 A.Reference number: A29769; PMID:87241248; PMID:3036070  
 A.Accession: A29769  
 A.Molecule type: protein  
 A.Residues: 18-166, 'X', 168-463 <ARL>  
 A:Note: 152-Leu was also found  
 R.Arlaud, G.J.; van Dorsselaer, A.; Bell, A.; Mancini, M.; Aude, C.; Gagnon, J.  
 Lett. 222, 129-134, 1987  
 A>Title: Identification of erythro-beta-hydroxyasparagine in the EGF-like domain of huma  
 A.Reference number: S02422; PMID:88005128; PMID:2820791  
 A.Accession: S02422  
 A.Molecule type: protein  
 A.Residues: 152-186 <AR3>  
 A:Note: 152-Leu was also found  
 R.Arlaud, G.J.; Gagnon, J.  
 Biochemistry 22, 1758-1764, 1983  
 A>Title: Complete amino acid sequence of the catalytic chain of human complement subcomp  
 A.Reference number: A00916; PMID:83204782; PMID:6303394  
 A.Accession: A00916  
 A.Molecule type: protein  
 A.Residues: 464-705 <AR2>  
 R.Thielems, N.M.; Aude, C.A.; Lacroix, M.B.; Gagnon, J.; Arlaud, G.J.  
 J. Biol. Chem. 265, 14469-14475, 1990  
 A>Title: Ca(2+) binding properties and Ca(2+)-dependent interactions of the isolated NH-  
 A.Reference number: A37820; PMID:90354439; PMID:2387866  
 A.Accession: A37820  
 A.Molecule type: protein  
 A.Residues: 18-26; 'V', 153-160; 'X', 252-255 <THI>  
 R.Pelloux, S.; Thielems, N.M.; Hudry-Clejeon, G.; Pettiloe, Y.; Filhol, O.; Arlaud, G.J.  
 FEBS Lett. 386, 15-20, 1996  
 A>Title: Identification of a cryptic protein kinase CK2 phosphorylation site in human CD

```

A:Reference number: S68830; MUID:96221263; PMID:8635594
A:Accession: S68830
A:Molecule type: protein
A:Residues: 133-137;187-211;610-613 <PEL>
A:Experimental source: plasma
C:Comment: C1r is a dimer of identical chains, each of which is activated by cleavage
n chain while fragment gamma remains disulfide bonded to the B chain to form C1r II.
C:Comment: This protein is a serine protease that combines with C1q and C1s to form C
n activate C2 and C4.
C:Genetics:
A:Gene: GDB:C1R
A:Cross-references: GDB:119729; OMIM:216950
A:Map position: 12p13-12p13
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto
C:keywords: acute phase; beta-hydroxyaspartate; calcium binding; complement pathway;
F:1-17/Domains: signal sequence #status predicted <SIG>
F:17-138/Domains: C1r/C1s repeat homology <C1R1>
F:18-463/Product: complement C1r chain A #status experimental <NCH>
F:146-189/Domains: EGF homology <EGF>
F:193-302/Domains: C1r/C1s repeat homology <C1R2>
F:297-463/Product: C1r gamma fragment #status experimental <GFR>
F:309-371/Domains: complement factor H repeat homology <FH1>
F:376-447/Domains: complement factor H repeat homology <FH2>
F:464-705/Product: complement C1r chain B #status experimental <BCB1>
F:464-697/Domains: trypsin homology <TRY>
F:71-89,146-165,161-174,176-189,193-220,250-268,309-358,338-371,376-429,406-447,451-5
F:123-221,514,581/Binding site: carbohydrate (asn) (covalent) #status experimental
F:167/Modified site: ephyr-beta-hydroxyaspartate (asn) #status experimental
F:206/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status experime
F:463-464/Cleavage site: Arg-Ile (autolytic) #status experimental
F:502,557,654/Active site: His, Asp, Ser #status predicted

Query Match          34.0%; Score 81; DB 1; Length 705;
Best Local Similarity 44.1%; Pred. No. 0.055;
Matches    15; Conservative     5; Mismatches   14; Indels      0; Gaps       0;
```

QY        8 PEPVFGRLASPGCEFYANDQERRWTLTAPGYR 41

```

Db              : 22 PQKLGCVTSPLFPKPYPNNFTVTIVTPNGR 55
                | : ||| : ||||| : ||| : ||| : |||
seizure-related protein SEZ-6 precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 31-Mar-2000
C:Accession: I52657
R:Shimizu-Nishikawa, K.; Kaijwara, K.; Kimura, M.; Katsuki, M.; Sugaya, E.
Brain Res. Mol. Brain Res. 28, 201-210, 1995
A>Title: Cloning and expression of SEZ-6, a brain-specific and seizure-related cDNA.
A:Reference number: I52657; MUID:95240392; PMID:773619
A:Accession: I52657
A>Status: preliminary; translated from GR/EMBL/DDBU
A:Molecule type: mRNA
A:Residues: 1-977 <RES>
A:Cross-references: GH:D29763; NID:g693909; PID:BAA06167.1; PID:g693910
C:Superfamily: complement factor H repeat homology
F:354-409/Domains: complement factor H repeat homology <FH01>
F:529-556/Domains: complement factor H repeat homology <FH02>
F:707-762/Domains: complement factor H repeat homology <FH03>
F:768-827/Domains: complement factor H repeat homology <FH04>
F:833-892/Domains: complement factor H repeat homology <FH05>

Query Match           33.8%; Score 80.5; DB 2; Length 977;
Best Local Similarity 36.5%; Pred. NO. 0.09;
Matches    19; Conservative     3; Mismatches   15; Indels     15; Caps      2;
```

QY        5 PKMV--PEPV-----FGRLSPCGPGEYANDQERRWTLTAPGYR 41

```

Db             400 PEMDSQDFVCIAACGCVAIRMTATGTGRVSPPGPNTSNNLTGMHLEPFESQR 451
```

RESULT 8

T09456  
Intrinsic factor-B12 receptor Cubilin precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 02-Aug-2002  
C:Accession: T09456  
R:Koziyaki, R.; Kristiansen, M.; Silahatoglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N.  
Blood 91, 3593-3600, 1998  
A:Title: The human intrinsic factor-vitamin B12 receptor, cubilin: Molecular characterization.  
A:Reference number: Z16677; MID:98241400; PMID:9572993  
A:Accession: T09456  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-3623 <KOZ>  
A:Cross-references: EMBL:AF034611; MID:g3929528; PIDN:AC82612.1; PID:g3929529  
C:Genetics:  
A:Map position: 10p12  
C:Superfamily: Intrinsic factor-B12 receptor cubilin; EGF homology  
C:Keywords: receptor; vitamin B12 uptake  
A:Keywords: signal sequence #status predicted <SIG>  
A:Keywords: Product: intrinsic factor-B12 receptor #status predicted <MAT>  
A:Keywords: Domain: EGF homology <EGF>  
A:Keywords: 467/Domain: EGF homology <EGF>  
Query Match 32.8%; Score 78; DB 2; Length 3623;  
Best Local Similarity 51.7%; Pred. No. 0.74;  
Matches 15; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
DB 12 FGRLASPGFGEYANDQERRMTLTAPPGY 40  
DB 3403 FGRLSPGWPNDNDKDTVTTLTAPNH 3431  
RESULT 9  
C1HUS  
complement subcomponent C 1SBR.GIF (EC 3.4.21.42) precursor [validated] - human  
N:Alternate names: C1 esterase precursor  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 03-Jun-2002  
C:Accession: A40496; A27381; S002732; S05634; A05140; A25396; A38407; B37820  
R:Kusumoto, H.; Hirose, S.; Sallier, J.P.; Hagen, F.S.; Kurachi, K.  
Proc. Natl. Acad. Sci. U.S.A. 85, 7307-7311, 1988  
A:Title: Human genes for complement components C1r and C1s in a close tail-to-tail array  
A:Reference number: A40496; MID:89017187; PMID:2459702  
A:Accession: A40496  
A:Molecule type: mRNA  
A:Residues: 1-688 <KUS>  
A:Cross-references: GB:J04080; MID:g179645; PIDN:AA51852.1; PID:g179646  
R:Rost, M.; Duponchel, C.; Meo, T.; Jullier, C.  
EMBL:J26, 8516-8524, 1987  
A:Title: Complete cDNA sequence of human complement C1s and close physical linkage of the  
A:Reference number: A27381; MID:88163522; PMID:2831944  
A:Accession: A27381  
A:Molecule type: mRNA  
A:Residues: 1-688 <ROS>  
A:Cross-references: GB:M18767; MID:g179647; PIDN:AA51853.1; PID:g179648  
R:MacKinnon, C.M.; Carter, P.E.; Smyth, S.J.; Dunbar, B.; Fothergill, J.E.  
Eur. J. Biochem. 169, 547-553, 1987  
A:Title: Molecular cloning of cDNA for human complement component C1s. The complete amir  
A:Reference number: S00224; MID:88082788; PMID:3500856  
A:Accession: S00224  
A:Molecule type: mRNA  
A:Residues: 1-688 <MAC>  
A:Cross-references: EMBL:X06596; MID:g29542; PIDN:CAA29817.1; PID:g763110  
A:Accession: S26732  
A:Molecule type: protein  
A:Residues: 16-36;66-116,170-236;246-262;265-280;282-284;287-308,315-363;384-394;421-435  
R:Tosi, M.; Duponchel, C.; Meo, T.; Couture-Tosi, E.  
J. Mol. Biol. 208, 709-714, 1989  
A:Title: Complement genes C1r and C1s feature an intronless serine protease domain close  
A:Reference number: S05634; MID:90040704; PMID:2553984  
A:Accession: S05634  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA

A:Residues: 356-513, 'G', 514-688 <RO2>  
R:Carter, P.E.; Dunbar, B.; Fothergill, J.E.  
Biochem. J. 215, 565-571, 1983  
A:Title: The serine protease chain of human complement component C1s. Cyanogen brom  
A:Reference number: A05140; MID:84104122; PMID:6362661  
A:Accession: A05140  
A:Molecule type: protein  
A:Residues: 438-483, 'X', 485-500;503-534;542-558;561-572, 'A', 574-601;617-623;626-644;6  
R:Syecher, S.E.; Nick, H.; Rickli, E.E.  
Eur. J. Biochem. 156, 49-57, 1986  
A:Title: Human complement component C1s. Partial sequence determination of the heavy  
A:Reference number: A25396; MID:86164350; PMID:3007145  
A:Accession: A25396  
A:Molecule type: protein  
A:Residues: 16-61;168-219;287-293, 'K', 295-334;384-445 <SPY>  
R:Hess, D.; Schaller, J.; Rickli, E.E.  
Biochemistry 30, 2827-2833, 1991  
A:Title: Identification of the disulfide bonds of human complement C1s.  
A:Reference number: A38407; MID:91175725; PMID:2007122  
A:Accession: A38407  
A:Molecule type: protein  
A:Residues: 131-134, 'X', 136-146, 'X', 148-150;155, 'X', 157-162;166-170, 'X', 172-174, 'X', 1  
'X', 387-402, 'X', 404-408;416-424, 'X', 426-431;547-556;592-597;617, 'X', 619-627, 'X', 629-  
R:Thieleens, N.M.; Audé, C.A.; Lacroix, M.B.; Gagnon, J.; Arlaud, G.J.  
J. Biol. Chem. 265, 14469-14475, 1990  
A:Title: Ca(2+) binding properties and Ca(2+)-dependent interactions of the isolated  
A:Reference number: A37820; MID:90354439; PMID:2387866  
A:Accession: B37820  
A:Molecule type: protein  
A:Residues: 16-25, 'X', 203-207 <THI>  
R:Thieleens, N.M.; Van Dorsselaer, A.; Gagnon, J.; Arlaud, G.J.  
Biochemistry 29, 3570-3578, 1990  
A:Title: Chemical and functional characterization of a fragment of C1s containing the  
A:Reference number: A32672; MID:90283368; PMID:2141278  
A:Contents: annotation; erythro-beta-hydroxyasparagine site, content  
A:Note: about half of the A chains contain erythro-beta-hydroxyasparagine  
C:Comment: This protein is a serine proteinase that combines with C1q and C1r to form  
C2 and C4.  
C:Comment: C1s is a dimer of identical chains, each of which is activated by cleavage  
C:Genetics:  
A:Gene: GDB:C1S  
A:Cross-references: GDB:119730; OMIM:120580  
A:Map position: 12p13-12p13  
A:Introns: 291/1; 329/3; 356/1; 399/1; 424/1  
A:Note: the list of introns may be incomplete  
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto  
C:Keywords: acute phase; beta-hydroxyasparagine; calcium binding; complement pathway;  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:11-127/Domain: C1r/C1s repeat homology <C1R1>  
F:16-688/Product: complement subcomponent C1s #status experimental <MAT>  
F:16-437/Product: complement subcomponent C1s chain A (heavy chain) #status experimen  
F:135-171/Domain: EGF homology <EGF>  
F:175-287/Domain: C1r/C1s repeat homology <C1R2>  
F:294-354/Domain: complement factor H repeat homology <FH1>  
F:359-421/Domain: complement factor H repeat homology <FH2>  
F:438-688/Product: complement subcomponent C1s chain B (light chain) #status experime  
F:438-675/Domain: trypsin homology <TRY>  
F:65-83;135-147;143-156;158-171;175-202;234-251;294-341;321-354;359-403;386-421;425-5  
F:149/Modified site: erythro-beta-hydroxyasparagine (Asn) (partial) #status experimen  
F:174;406/Binding site: carbonylde (Asn) (covalent) #status experimental  
F:437-438/Cleavage site: Arg-116 (complement subcomponent C1r) #status experimental  
F:475;529;632/Active site: His, Asp, Ser #status predicted  
Query Match 31.3%; Score 74.5; DB 1; Length 688;  
Best Local Similarity 34.3%; Pred. No. 0.33; Mismatches 13; Indels 1; Gaps 1;  
Matches 12; Conservative 9;  
DB 7 WPEP-VGRLASPGFGEYANDQERRMTLTAPPGY 40  
DB 14 YAEPTMGTGELTSPYDPAVSEVEKSMIDIEVPGY 48  
RESULT 10



A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1070 <LID>  
 A:Cross-references: EMBL:U57369; NID:g1899041; PID:g1899042; PIDN:AAC47485.1  
 C:Superfamily: dorsal-ventral patterning protein tollloid; astacin homology; C1r/C1s repeat

Query Match 30.0%; Score 71.5; DB 2; Length 1070;  
 Best Local Similarity 39.5%; Pred. No. 1.2;  
 Matches 15; Conservative 5; Mismatches 17; Indels 1; Gaps 1;

OY 4 GPKMPEVFGRLSPGPEYANDQERRMTLTAPPGR 41  
 Db 521 GGNNPGE-GFLNSPAYDEYSDKVCWITVREGYQ 557

## RESULT 15

BMH1  
 procollagen C-endopeptidase (EC 3.4.24.19) precursor, splice form BMP1 - human  
 N:Alternate names: bone morphogenic protein 1 (BMP1)

A:Species: Homo sapiens (man)  
 A:Accession: A37278; E58788  
 A:Title: Novel regulators of bone formation: molecular clones and activities.

A:Reference number: A37278; M0ID:89072730; PMID:3201241  
 A:Accession: A37278  
 A:Molecule type: mRNA

A:Residues: 1-730 <WO2>  
 A:Cross-references: GB:M22488; NID:g179499; PIDN:AAA51833.1; PID:g179500  
 C:Genetics:

A:Gene: GDB:BMP1  
 A:Cross-references: GDB:125203; OMIM:112264  
 A:Map position: 8p21-8p21

## C:Function:

A:Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen type C:Superfamily: procollagen C-endopeptidase; astacin homology; C1r/C1s repeat homology; E C:Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication; g1 F:1-22/Domain: signal sequence #status predicted <Sig>  
 F:23-730/Product: procollagen C-endopeptidase splice form BMP1 #status predicted <Mat>  
 F:130-321/Domain: astacin homology <AST>  
 F:322-431/Domain: C1r/C1s repeat homology <C1R1>  
 F:435-544/Domain: C1r/C1s repeat homology <C1R2>  
 F:551-587/Domain: EGF homology <EGF>

F:591-700/Domain: C1r/C1s repeat homology <C1R3>  
 F:91,142,332,363,599/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:163-319,185-205,322-348,375-397,435-461,488-510,551-563,559-572,574-587,591-617,644-66 F:213,217,223,272/Binding site: zinc (His, His, Tyr) #status predicted  
 A:Title: Novel regulators of bone formation: molecular clones and activities.  
 A:Reference number: A37278; M0ID:89072730; PMID:3201241  
 A:Accession: A37278  
 A:Molecule type: mRNA  
 A:Residues: 1-730 <WO2>  
 A:Cross-references: GB:M22488; NID:g179499; PIDN:AAA51833.1; PID:g179500  
 C:Genetics:

Query Match 29.0%; Score 69; DB 1; Length 730;  
 Best Local Similarity 44.8%; Pred. No. 1.6;  
 Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

OY 13 GRLASPGFPEYANDQERRMTLTAPPGR 41  
 Db 600 GSITSPGMPREYPPNKNKCIQVAPPTQYR 628

## RESULT 16

A58788  
 procollagen C-endopeptidase (EC 3.4.24.19) precursor, splice form HIS - human  
 N:Alternate names: bone morphogenic protein splice form BMP-1/HIS

C:Species: Homo sapiens (man)  
 C:Date: 28-Mar-1998 #sequence\_revision 09-Apr-1998 #text\_change 18-Jun-1999  
 C:Accession: A37278; A58788  
 R:Mooney, J.M.; Rosen, V.; Celeste, A.J.; Mltsock, L.M.; Whitters, M.J.; Kriz, R.W.; Hew Science 242, 1528-1534, 1998

A:Title: Novel regulators of bone formation: molecular clones and activities.  
 A:Reference number: A37278; M0ID:89072730; PMID:3201241  
 A:Accession: A37278  
 A:Molecule type: mRNA

A:Residues: 1-702, 'EKRPALPPRGPRHQLKFRVQKRNKTPQ' <WO2>  
 A:Cross-references: GB:M22488; NID:g179499; PIDN:AAA51833.1; PID:g179500  
 R:Takahara, K.; Lyons, G.E.; Greenspan, D.S.  
 J. Biol. Chem. 269, 32572-32578, 1994

A:Title: Bone morphogenetic protein-1 and a mammalian tollloid homologue (mTld) are en A:Reference number: A58788; M0ID:95096114; PMID:7798260  
 A:Molecule type: mRNA  
 A:Accession: A58788  
 A:Residues: 703-823 <TAK>

A:Cross-references: GB:L35278; NID:g619423; PIDN:AAC41703.1; PID:g619424  
 C:Genetics:

A:Gene: GDB:BMP1; BMP-1  
 A:Cross-references: GDB:125203; OMIM:112264  
 A:Map position: 8p21-8p21

## C:Function:

A:Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen t C:Superfamily: procollagen C-endopeptidase; astacin homology; C1r/C1s repeat homology C:Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication; F:1-22/Domain: signal sequence #status predicted <Sig>  
 F:23-823/Product: procollagen C-endopeptidase splice form HIS #status predicted <Mat>  
 F:130-321/Domain: astacin homology <AST>  
 F:322-431/Domain: C1r/C1s repeat homology <C1R1>  
 F:435-544/Domain: C1r/C1s repeat homology <C1R2>  
 F:551-587/Domain: EGF homology <EGF>

F:591-700/Domain: C1r/C1s repeat homology <C1R3>  
 F:738-752/Region: histidine-rich  
 F:91,142,332,363,599/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:163-319,185-205,322-348,375-397,435-461,488-510,551-563,559-572,574-587,591-617,644 F:213,217,223,272/Binding site: zinc (His, His, Tyr) #status predicted  
 F:214/Active site: Glu #status predicted  
 F:565/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 29.0%; Score 69; DB 1; Length 823;  
 Best Local Similarity 44.8%; Pred. No. 1.9;  
 Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

OY 13 GRLASPGFPEYANDQERRMTLTAPPGR 41  
 Db 600 GSITSPGMPREYPPNKNKCIQVAPPTQYR 628

## RESULT 17

B58788  
 procollagen C-endopeptidase (EC 3.4.24.19) precursor, tollloid-like splice form - huma N:Alternate names: bone morphogenic protein 1, tollloid-like splice form

C:Species: Homo sapiens (man)  
 C:Date: 28-Mar-1998 #sequence\_revision 09-Apr-1998 #text\_change 16-Jul-1999  
 C:Accession: A37278; B58788  
 R:Mooney, J.M.; Rosen, V.; Celeste, A.J.; Mltsock, L.M.; Whitters, M.J.; Kriz, R.W.; Science 242, 1528-1534, 1998

A:Title: Novel regulators of bone formation: molecular clones and activities.  
 A:Reference number: A37278; M0ID:89072730; PMID:3201241  
 A:Accession: A37278  
 A:Molecule type: mRNA

A:Residues: 1-702, 'EKRPALPPRGPRHQLKFRVQKRNKTPQ' <WO2>  
 A:Cross-references: GB:M22488; NID:g179499; PIDN:AAA51833.1; PID:g179500  
 R:Takahara, K.; Lyons, G.E.; Greenspan, D.S.  
 J. Biol. Chem. 269, 32572-32578, 1994

A:Title: Bone morphogenetic protein-1 and a mammalian tollloid homologue (mTld) are en A:Reference number: A58788; M0ID:95096114; PMID:7798260  
 A:Molecule type: mRNA  
 A:Accession: B58788  
 A:Residues: 703-986 <TAK>

A:Cross-references: GB:L35279; NID:g619860; PIDN:AAC41710.1; PID:g619861  
 C:Genetics:

A:Gene: GDB:BMP1; BMP-1  
 A:Cross-references: GDB:125203; OMIM:112264  
 A:Map position: 8p21-8p21

A:Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen t C:Superfamily: procollagen C-endopeptidase; astacin homology; C1r/C1s repeat homology C:Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication; F:1-22/Domain: signal sequence #status predicted <Sig>





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:42:59 : Search time 121 Seconds  
(without alignments)  
69.818 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238  
Sequence: 1 TPLGPKWPEPVFGRLASPGF.....GEYANDQERRWTLTAPGVR 41

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

ed: 671580 segs, 206047115 residues

number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :

SPREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	238	100.0	175	4 Q9UMV3	Q9umv3 homo sapien
2	228	100.0	181	4 Q9UDC7	Q9udc7 homo sapien
3	238	100.0	185	4 Q9UBP3	Q9ubp3 homo sapien
4	183	76.9	185	11 Q9QX44	Q9qxd4 mus musculu
5	183	76.9	190	11 Q9QX05	Q9qxd5 mus musculu
6	183	76.9	367	11 Q9QXD2	Q9qxd2 mus musculu
7	183	76.9	685	11 Q9Z338	Q9z338 mus musculu
8	183	76.9	685	11 Q91WP0	Q91wp0 mus musculu
9	180	75.6	678	11 Q9JJS8	Q9jjs8 rattus norv
10	170	71.4	186	11 Q9QXD4	Q9qxd4 rattus norv
11	164	68.9	177	11 Q9QX86	Q9qxd6 rattus norv
12	164	68.9	185	11 Q9QX88	Q9qxd8 rattus norv
13	164	68.9	216	11 Q9QX83	Q9qxd3 rattus norv
14	100	42.0	688	13 Q9PV74	Q9pvy4 xenopus lae
15	93	39.1	212	11 Q9Q020	Q9q020 rattus norv
16	93	39.1	701	11 Q9JJS9	Q9jjs9 rattus norv

Hit No. 3 has  
best biological interest  
including in the  
context of this  
panel 2

## ALIGNMENTS

17	93	39.1	728	4 Q96RS4	Q96rs4 homo sapien
18	93	39.1	733	11 Q920S0	Q920s0 mus musculu
19	91.5	38.4	996	4 Q8TD25	Q8td25 homo sapien
20	87	36.6	16	4 Q9UC48	Q9uc48 homo sapien
21	84	35.3	3138	5 Q9VTP0	Q9vtp0 drosophila
22	83.5	35.1	384	13 Q9W630	Q9w630 cyprinus ca
23	83.5	35.1	1290	11 Q62827	Q62827 rattus norv
24	83	34.9	719	13 Q9PVY2	Q9pvy2 triakis scy
25	81.5	34.2	745	13 Q9PVY3	Q9pvy3 cyprinus ca
26	80.5	33.8	605	11 Q62224	Q62224 mus musculu
27	80.5	33.8	977	11 Q62269	Q62269 mus musculu
28	80.5	33.8	991	11 Q62223	Q62223 mus musculu
29	80	33.6	832	11 Q9JLB4	Q9jlb4 mus musculu
30	78	32.8	1048	4 Q96P27	Q96p27 homo sapien
31	78	32.8	3389	4 Q96P29	Q96p29 homo sapien
32	78	32.8	3508	4 Q96RM4	Q96rm4 homo sapien
33	78	32.8	3564	11 Q9Z3L3	Q9z3l3 mus musculu
34	78	32.8	3623	4 Q6Q494	Q6q494 homo sapien
35	78	32.8	4215	5 Q9W332	Q9w332 drosophila
36	76	31.9	722	13 Q9PS25	Q9ps25 lampetra ja
37	76	31.9	3620	6 Q9TU53	Q9tu53 canis famil
38	75	31.5	421	11 Q89002	Q89002 mus musculu
39	75	31.5	707	11 Q9ET60	Q9et60 mus musculu
40	75	31.5	707	11 Q99K16	Q99k16 mus sapien
41	74.5	31.3	139	4 Q96G02	Q96g02 homo sapien
42	73.5	30.9	273	5 Q97070	Q97070 ascaris suu
43	73.5	30.9	288	5 Q9U994	Q9u994 ascaris suu
44	73	30.7	107	4 Q9H4W1	Q9h4w1 homo sapien
45	73	30.7	170	4 Q9H4W2	Q9h4w2 homo sapien
46	72.5	30.5	686	13 Q9DGC2	Q9dgc2 cyprinus ca
47	72	30.3	3623	11 Q70244	Q70244 rattus norv
48	71.5	30.0	414	11 Q9CX06	Q9cx06 mus musculu
49	71.5	30.0	414	11 Q8R4W6	Q8r4w6 mus musculu
50	71.5	30.0	1070	5 P91972	P91972 aplysia cal
51	70.5	29.6	685	13 Q9DGC1	Q9dgc1 cyprinus ca
52	70.5	29.6	685	13 Q9DGC0	Q9dgc0 cyprinus ca
53	70	29.4	698	13 Q9PU71	Q9pu71 xenopus lae
54	70	29.4	997	11 Q9JUS0	Q9jus0 mus musculu
55	69	29.0	241	11 Q9Z135	Q9z135 rattus norv
56	69	29.0	735	13 Q57381	Q57381 xenopus lae
57	69	29.0	999	4 Q9NQ36	Q9nq36 homo sapien
58	69	29.0	1464	5 Q24132	Q24132 drosophila
59	69	29.0	1464	5 Q23995	Q23995 drosophila
60	69	29.0	1464	5 Q9VC47	Q9vc47 drosophila

## RESULT 1

Q9UMV3	PRELIMINARY:	PRT:	175 AA.
AC Q9UMV3:			
DT 01-MAY-2000 (TREMBLrel. 13, Created)			
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE Manose binding lectin-associated serine protease-2 related protein,			
DE MASP-2.			
GN Homo sapiens (Human).			
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX NCBI-TaxID=9606;			
RN (1)			
RP SEQUENCE FROM N.A.			
RC TISSUE=LIVER;			
RA MEDLINE=99192764; PubMed=10092804;			
RA Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,			
RA Jensenius J.C., Schwaebble W.J.;			
RT Two constituents of the initiation complex of the manose-binding			
RT lectin activation pathway of complement are encoded by a single			
RT structural gene.";			
RL J. Immunol. 162:3481-3490(1999).			



[illegible]

OS	MAP19.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6 X CBA;
RX	MEDLINE=20054576; PubMed=10586086;
RA	Stover C.M., Thiel S., Lynch N.J., Schwaeble W.J.;
RT	"The rat and mouse homologs of MASP-2 and MAP19, components of the
RT	mamnan-binding lectin activation pathway of complement.";
RL	J. Immunol. 163:6848-6859(1999).
CC	-I- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR	EMBL; Y19160; CAB65247.1; -.
DR	HSSP; P00736; IAPQ.
DR	MEROPS; S01.229; -.
DR	InterPro; IPRO00152; Asx_hydroxyl.
DR	InterPro; IPRO00859; CUB_domain.
DR	InterPro; IPRO00561; EGF-like.
DR	InterPro; IPRO01881; EGF_Ca.
DR	Pfam; PF00431; CUB; 1.
DR	SMART; SM00042; CUB; 1.
DR	SMART; SM00179; EGF_CA; 1.
DR	PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR	PROSITE; PS01180; CUB; 1.
DR	PROSITE; PS01186; EGF_2; 1.
KW	PROSITE; PS01187; EGF_CA; 1.
KW	Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease;
KW	Repeat; Signal.
FT	SIGNAL 1
SO	SEQUENCE 190 AA; 21457 MW; 22BCBAA3F905B010 CRC64;
SC	POTENTIAL.
Query Match	76.9%; Score 183; DB 11; Length 190;
Best Local Similarity	78.0%; Pred. NO. 4.9e-16;
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0.	
Dy	1 TPLGPKWEPEVFGRLASPGFGEVANDERRMTLTAPPGR 41                     :  :               Db 21 TLGSKWEPEVGRLVSPGFEPKYADHDRSMTLTAPPGR 61
RESULT 6	
Q9QXD2	
ID	Q9QXD2 PRELIMINARY; PRT; 367 AA.
AC	Q9QXD2.
DT	01-MAY-2000 (TREMBLrel. 13. Created)
DT	01-MAY-2000 (TREMBLrel. 13. Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21. Last annotation update)
DE	Mannose binding lectin-associated serine protease-2 precursor (Fragment).
DE	MASP2 OR MASP-2.
GN	Mus musculus (Mouse).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=B6CBAF1/J.
RX	MEDLINE=20054576; PubMed=10586086;
RA	Stover C.M., Thiel S., Lynch N.J., Schwaeble W.J.;
RT	"The rat and mouse homologs of MASP-2 and MAP19, components of the
RT	mamnan-binding lectin activation pathway of complement.";
RL	J. Immunol. 163:6848-6859(1999).
CC	-I- SUBCELLULAR LOCATION: SECRETED (BY SIMILARTY).
CC	-I- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR	EMBL; Y19163; CAB65250.1; -.
DR	HSSP; P00736; IAPQ.
DR	MEROPS; S01.229; -.
DR	MCD; MGI:I330833; Masp2.
DR	InterPro; IPRO00152; Asx_hydroxyl.
DR	InterPro; IPRO00859; CUB_domain.

DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam: PF000431; CUB; 2.  
 DR Pfam: PF00084; sushi; 1.  
 DR PRINTS: PR00277; INSULIN.  
 DR SMART: SM00032; CCP; 1.  
 DR SMART: SM00042; CUB; 2.  
 DR SMART: SM00179; EGF\_CA; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_CA; 1.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease;  
 KW Repeat; signal;  
 FT SIGNAL 1 20 POTENTIAL.  
 FT NON\_TER 367 367  
 SQ SEQUENCE 367 AA; 40950 MW; F66FCB29CEAB2B7D CRC64;

Query Match 76.9%; Score 183; DB 11; Length 367;  
 Local Similarity 78.0%; Pred. No. 1.1e-15;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 TPLGPKWPEVFGRLASPGFPGGEYANDQERRMTLTAPPGYR 41  
 Db 21 TLGSKWPEVFGRLVSPGFPEKYADHDQRSWTLTAPPGYR 61

RESULT 7  
 ID 092338 PRELIMINARY: PRT; 685 AA.  
 AC 092338;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Manose-binding lectin associated serine protease-2.  
 GN MASP2 OR MASP-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-LIVER;  
 RX MEDLINE=99008558; PubMed=9794427;  
 RA Endo Y., Takahashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,  
 Nonaka M., Fujita T.;  
 ABSTRACT: Two lineages of manose-binding lectin-associated serine protease  
 (MASP) in vertebrates.";  
 J. Immunol. 161:4924-4930(1998).  
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPSIN FAMILY.  
 CC EMBL; AB009459; BAA34674.1; -  
 DR HSSP; P00763; IDPO.  
 DR MEROPS: S01.229; -  
 DR MGD: MGT.1330832; Masp2.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR000859; CUB\_domain.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam: PF00431; CUB; 2.  
 DR Pfam: PF00084; sushi; 2.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00032; CCP; 2.  
 DR SMART: SM00042; CUB; 2.  
 DR SMART: SM00179; EGF\_CA; 1.  
 DR SMART: SM00020; Tryp\_Spc; 1.

DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_CA; 1.  
 DR PROSITE: PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Lectin;  
 KW protease; Repeat; Serine protease.  
 KW SEQUENCE 685 AA; 75490 MW; 4571E1AE0703A70F CRC64;

Query Match 76.9%; Score 183; DB 11; Length 685;  
 Best Local Similarity 78.0%; Pred. No. 2.2e-15;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 TPLGPKWPEVFGRLASPGFPGGEYANDQERRMTLTAPPGYR 41  
 Db 16 TLGSKWPEVFGRLVSPGFPEKYADHDQRSWTLTAPPGYR 56

RESULT 8  
 ID 091WPO PRELIMINARY: PRT; 685 AA.  
 AC 091WPO;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Similar to mannan-binding lectin serine protease 2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RA Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC013893; AAH13893.1; -  
 DR MEROPS: S01.229; -  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000859; CUB\_domain.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam: PF00431; CUB; 2.  
 DR Pfam: PF00084; sushi; 2.  
 DR Pfam: PF00089; trypsin; 1.  
 DR SMART: SM00181; EGF; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE: PS01187; EGF\_CA; UNKNOWN\_1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS50240; TRYPSIN\_SER; 1.  
 DR PROSITE: PS00135; TRYPSIN\_DOM; UNKNOWN\_1.  
 DR Hydrolyase; Lectin; Protease; Serine protease.  
 KW SEQUENCE 685 AA; 75516 MW; F56A6D52BC7099D CRC64;

Query Match 76.9%; Score 183; DB 11; Length 685;  
 Best Local Similarity 78.0%; Pred. No. 2.2e-15;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 TPLGPKWPEVFGRLASPGFPGGEYANDQERRMTLTAPPGYR 41  
 Db 16 TLGSKWPEVFGRLVSPGFPEKYADHDQRSWTLTAPPGYR 56

RESULT 9  
 ID 09JUS8 PRELIMINARY: PRT; 678 AA.  
 AC 09JUS8;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)







```
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00084; Sush1; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Protease; Repeat; Serine protease.
SEQUENCE 688 AA; 77513 MW; 180DD559EF8B6573 CRC64;

Query Match
Best Local Similarity 42.0%; Score 100; DB 13; Length 688;
Matches 16; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 11 VFGRLASPGFPGEVANDQERRWTLTAPPGYR 41
DB 29 LFGRISSPGFPRKPSYSDLTMMNINIVPEGR 59

RESULT 15
009020 PRELIMINARY; PRT; 212 AA.
ID: 009020
AC: 009020;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ra-reactive factor serine protease p100 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuitel T., Fellmer P., Neubauer K., Kawakami M., Grundmann A.,
RA Ramadori G.;
RT "The complement activating protease p100 is expressed by hepatocytes
nd is induced by IL-6 in vitro and during the acute phase reaction in
lvo.";
ab. Invest. 0:0-0(1997).
I-SIMILARITY: CONTAINS 1 CUB DOMAIN.
EMBL: AF004661; AAB65832.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.198; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR Pfam: PF00431; CUB; 2.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Protease; Repeat.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 24377 MW; 1BB2F52BF0D30A7E CRC64;

Query Match
Best Local Similarity 39.1%; Score 93; DB 11; Length 212;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 11 VFGRLASPGFPGEVANDQERRWTLTAPPGYR 41
DB 21 MFCQIOSPGYPSDSEVTWITVPEGR 51

RESULT 16
09JJS9 PRELIMINARY; PRT; 701 AA.
ID 09JJS9
AC 09JJS9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannose-binding protein associated serine protease-1 precursor
DE (Fragment).
GN MASP-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC MEDLINE=20469449; PubMed=10913141;
RA Wallis R., Dodd R.B.;
RT "Interaction of mannose-binding protein with associated serine
proteases: Effects of naturally occurring mutations.";
RL J. Biol. Chem. 275:30962-30969(2000).
CC -I-SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -I-SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: AJ277423; CAB89695.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.198; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; Sush1; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Protease;
KW Repeat; Serine protease; Signal.
FT NON_TER 1
FT SIGNAL <1 21
FT CHAIN 22 701
FT CHAIN 701
SQ SEQUENCE 701 AA; 79663 MW; B5CFD619D63CE3DD CRC64;

Query Match
Best Local Similarity 39.1%; Score 93; DB 11; Length 701;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 11 VFGRLASPGFPGEVANDQERRWTLTAPPGYR 41
DB 29 MFCQIOSPGYPSDSEVTWITVPEGR 59

RESULT 17
```

Q96RS4 ID Q96RS4 PRELIMINARY: PRT: 728 AA.

AC Q96RS4: 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Complement factor MASP-3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxId=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21378425; PubMed=11485744;

RA Dahl M.R., Thiel S., Matsushita M., Fujita T., Willis A.C., Christensen T., Vorup-Jensen T., Jensenius J.C.;

RT "Maspl-3 and its association with distinct complexes of the mann- binding lectin complement activation pathway.";

RL Immunity 15:127-135(2001).

KW EMBL; AF284421; AAK84071.1; -.

DR InterPro: IPR000859; CUB\_domain.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001881; EGF\_Ca.

DR InterPro: IPR001254; Ser\_protease\_Try.

DR Pfam: PF00431; CUB; 2.

DR Pfam: PF00084; sushi; 2.

DR Pfam: PF00089; trypsin; 1.

DR SMART; SM00181; EGF; 1.

DR PROSITE; PS001010; ASX\_HYDROXYL; UNKNOWN\_1.

DR PROSITE; PS01180; CUB; 2.

DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE; PS01187; EGF\_CA; UNKNOWN\_1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.

DR Hydrolase; Serine protease.

SO SEQUENCE 728 AA; 81860 MW; 09B5297A6C14283A CRC64;

Query Match 39.1%; Score 93; DB 4; Length 728;

Best Local Similarity 45.2%; Pred. No. 0.0012;

Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 11 VGRLASPGFPGFYANDERRWTLTAPGGR 41

DB 27 MEGQIQSPGYPDPSPDSEVTWNITVPEGFR 57

18

92050 PRELIMINARY: PRT: 733 AA.

AC Q92050: 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE MBL-associated serine protease-3.

GN MASP1 OR MASP3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxId=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-BALB/C; TISSUE=LIVER;

RA Takahashi M., Miura S., Ishii N., Sugamura K., Shuichi M., Shiro S., Endo Y., Matsushita M., Fujita T.;

RT "An essential role of MBL-associated serine protease-1 and -3 in activation of complement by lectin pathway.";

RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL; AB049755; BAB69688.1; -.

DR MEROPS; S01.132; -.

DR MGD; MGI:88492; Masp1.

DR InterPro: IPR000152; Asx\_hydroxyl.

DR InterPro: IPR000859; CUB\_domain.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001881; EGF\_Ca.

DR InterPro: IPR001254; Ser\_protease\_Try.

DR InterPro: IPR000436; Sushi\_SCR\_CCP.

DR Pfam; PF00431; CUB; 2.

DR Pfam; PF00084; sushi; 2.

DR Pfam; PF00089; trypsin; 1.

DR SMART; SM00181; EGF; 1.

DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_1.

DR PROSITE; PS01180; CUB; 2.

DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE; PS01187; EGF\_CA; UNKNOWN\_1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.

KW Hydrolase; Protease; Serine protease.

SO SEQUENCE 733 AA; 82373 MW; 6B917743118941C1 CRC64;

Query Match 39.1%; Score 93; DB 11; Length 733;

Best Local Similarity 45.2%; Pred. No. 0.0012;

Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 11 VGRLASPGFPGFYANDERRWTLTAPGGR 41

DB 32 MEGQIQSPGYPDPSPDSEVTWNITVPEGFR 62

19

08TD25 PRELIMINARY: PRT: 996 AA.

AC Q8TD25: 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE SE26.

GN SE26.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

RN NCBI\_TaxId=9606;

RP SEQUENCE FROM N.A.

RA Yu L.;

RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AY038048; AAK71497.1; -.

SO SEQUENCE 996 AA; 107585 MW; E8012D30969925E CRC64;

Query Match 38.4%; Score 91.5; DB 4; Length 996;

Best Local Similarity 40.4%; Pred. No. 0.0028;

Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

QY 5 PKW-PEPY-----FGRLASPGFPGFYANDERRWTLTAPGGR 41

DB 405 PWDSEKPEVCIECPGVIKATGRIIVSPFGNYNNLTCHWLLDAPGGR 456

20

09UC48 PRELIMINARY: PRT: 16 AA.

AC Q9UC48: 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)

DE 18.5 kDa tumor marker protein (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxId=9606;

RN [1]

RP SEQUENCE

RX MEDLINE=96212543; PubMed=8618346;

RA Rasmussen H.H., Ornoff T.F., Wolf H., Celis J.E.;

RT "Towards a comprehensive database of proteins from the urine of

RT patients with bladder cancer."  
 RL J. Urol. 155:2113-2119(1996).  
 SQ SEQUENCE 16 AA; 1751 MM; 289540537A1B4400 CRC64;

Query Match 36.6%; Score 87; DB 4; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LASPGFPGGYANDQER 30  
 ||||||||||||  
 Db 1 LASPGFPGGYANDQER 16

Search completed: January 11, 2003, 10:51:43  
 Job time : 125 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:08:28 ; Search time 65 seconds

(without alignments)  
26.162 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPLGPKWPEVFGRLASPGF.....GEVANDERRWLTAPPGRY 41

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ed: 112892 seqs, 41476328 residues

number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	686	1 MAS2_HUMAN	000187 homo sapien
2	93	39.1	699	1 CRAR_HUMAN	P48740 h complemen
3	93	39.1	704	1 CRAR_MOUSE	P98064 mus musculu
4	85	35.7	592	1 TOR2_CAEEL	P98060 caenorhabdi
5	81	34.0	705	1 CLR_HUMAN	P00736 homo sapien
6	78	32.8	449	1 PCOI_HUMAN	015113 homo sapien
7	74.5	31.3	688	1 C1S_HUMAN	P09871 homo sapien
8	74	31.1	468	1 PCOI_MOUSE	061398 mus musculu
9	74	31.1	514	1 UVS2_XENLA	008628 rattus norv
10	74	31.1	1022	1 TLID_BRARE	057460 brachydiano
11	74	31.1	695	1 CASP_MESAU	P15156 mesocricetu
12	72.5	30.5	986	1 BMP1_HUMAN	P13447 homo sapien
13	69	29.0	991	1 BMP1_MOUSE	P98063 mus musculu
14	69	29.0	707	1 BMP1_XENLA	P98070 xenopus lae
15	67	28.2	639	1 BMP1_STRPU	P98069 strongyloce
16	66	27.7	914	1 NRPI_CHICK	P79795 gallus gall
17	66	27.5	503	1 CA19_CHICK	P12106 gallus gall
18	65.5	27.3	1057	1 TLID_DROME	P25723 drosophila
19	65	26.9	925	1 NRPI_RAT	035276 rattus norv
20	64	26.9	922	1 NRPI_RAT	09qwl9 rattus norv
21	63	26.5	923	1 NRPI_HUMAN	014766 homo sapien
22	63	26.5	923	1 NRPI_MOUSE	P97333 mus musculu
23	63	26.5	931	1 NRPI_HUMAN	060462 homo sapien
24	63	26.5	931	1 NRPI_MOUSE	P02462 homo sapien
25	63	26.5	931	1 NRPI_HUMAN	035375 mus musculu
26	61	25.6	931	1 NRPI_MOUSE	P02462 homo sapien
27	60.5	25.4	1262	1 CA13_CHICK	P12105 gallus gall
28	60	25.2	1466	1 CA13_HUMAN	P02461 homo sapien
29	59.5	25.0	1024	1 SZ6L_HUMAN	09bvl1 homo sapien
30	59	24.8	636	1 CA13_RAT	P13911 rattus norv
31	58	24.4	1464	1 CA13_MOUSE	P08121 mus musculu
32	58	24.4	1670	1 CA13_HUMAN	P01955 homo sapien
33	57.5	24.2	275	1 TSG6_MOUSE	008859 mus musculu

Hit No. 1 has bidio-  
gous play component  
to hit no. 3 and 33  
results.

## ALIGNMENTS

RESULT 1	STANDARD:	PRT:	686 AA.
ID MAS2_HUMAN	000187: 075754; Q9Y270; Q9B2H0;		
AC 30-MAY-2000 (Rel. 39, Created)			
DT 30-MAY-2000 (Rel. 39, Last sequence update)			
DT 15-JUN-2002 (Rel. 41, Last annotation update)			
DE Mannan-binding lectin serine protease 2 precursor (EC 3.4.21.-)			
DE (Mannose-binding lectin protein associated serine protease 2) (MASP-2)			
GN MASP2.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Liver;			
RX MEDLINE-97242412; PubMed-9087411;			
RA Thiel S., Jensen T.V., Stover C.M., Schwaible W.J., Laursen S.B.,			
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,			
RA Reid K.B.M., Jensenius J.C.;			
RT "A second serine protease associated with mannan-binding lectin that			
RT activates complement";			
RL Nature 386:506-510(1997).			
RN [2]			
RP SEQUENCE FROM N.A.			
RA Thiel S., Vorup-Jensen T., Stover C.M., Schwaible W., Laursen S.B.,			
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,			
RA Reid K.B.M., Jensenius J.C.;			
RT "Identification and characterization of a novel protein of the human			
RT complement system, mannan-binding lectin-associated serine protease-2			
RT (MASP-2).";			
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.			
RN [3]			
RP SEQUENCE FROM N.A.			
RX MEDLINE-99192764; PubMed-10092804;			
RA Stover C.M., Thiel S., Theisen M., Lynch N.J., Vorup-Jensen T.,			
RA Jensenius J.C., Schwaible W.J.;			
RT "Two constituents of the initiation complex of the mannan-binding			
RT lectin activation pathway of complement are encoded by a single			

34	57.5	24.2	277	1 TSG6_HUMAN	P98066 homo sapien
35	57.5	24.2	1685	1 CAS4_HUMAN	P29400 homo sapien
36	57.5	24.2	2319	1 NTC3_RAT	09r172 rattus norv
37	57	23.9	822	1 PBPA_PSEAE	007806 pseudomonas
38	56.5	23.7	276	1 TSG6_RABIT	P98065 oryctolagus
39	56.5	23.7	1049	1 CA13_BOVIN	P04238 bos taurus
40	56.5	23.7	1376	1 VCAP_HSEVB	P28920 equine herp
41	56	23.5	597	1 NCPI_PARLI	P42674 paracentrot
42	56	23.5	616	1 SPAN_STRPU	P98068 strongyloce
43	56	23.5	1669	1 CA14_MOUSE	P02453 mus musculu
44	55.5	23.3	1690	1 CA44_HUMAN	P53440 homo sapien
45	55	23.1	241	1 LEQ3_RABIT	P47845 oryctolagus
46	55	23.1	283	1 YQ33_CAEEL	009233 caenorhabdi
47	54.5	22.9	323	1 CC39_CAEEL	009455 caenorhabdi
48	54.5	22.9	1496	1 CA25_HUMAN	P05997 homo sapien
49	54.5	22.9	2321	1 NTC3_HUMAN	08um47 homo sapien
50	54	22.7	411	1 PEPT_YERPE	082t70 yeastlike pe
51	54	22.7	443	1 COBB_THEAC	09h1x6 thermoplasm
52	54	22.7	777	1 PBPB_VIBCH	Q9KUC0 v penicillini
53	54	22.7	928	1 NRPI_XENLA	P28824 xenopus lae
54	54	22.7	1174	1 KPCI_COCHF	042632 cochlinoctu
55	53.5	22.5	476	1 MM10_MOUSE	055123 mus musculu
56	53.5	22.5	680	1 CA1A_MOUSE	005306 mus musculu
57	53.5	22.5	891	1 SYV_PYRHO	058052 pyrococcus
58	53.5	22.5	1328	1 YMT5_YEAST	004214 saccharomyc
59	53.5	22.5	1328	1 YMT5_YEAST	004670 saccharomyc
60	53.5	22.5	1755	1 YJ29_YEAST	P47100 saccharomyc

RT structural gene.";  
 RL J. Immunol. 162:3481-3490(1999).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Park D., Kim B., Baek K., Yoon J.;  
 "Structure of human MASP-2 gene.";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: TRYPSIN PROTEASE THAT PRESUMABLY PLAYS AN IMPORTANT ROLE  
 IN THE INITIATION OF THE MBL COMPLEMENT ACTIVATION PATHWAY. AFTER  
 ACTIVATION IT CLEAVES C4 GENERATING C4a AND C4b.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
 CC -----  
 DR EMBL: Y09926; CAA71059.1; -  
 DR EMBL: X98400; CAA67050.1; -  
 DR EMBL: Y18287; CAB50735.1; -  
 DR EMBL: Y18286; CAB50733.1; -  
 DR EMBL: AF321562; AAG50274.1; -  
 DR EMBL: AF321558; AAG50274.1; JOINED.  
 DR EMBL: AF321559; AAG50274.1; JOINED.  
 DR EMBL: AF321560; AAG50274.1; JOINED.  
 DR EMBL: AF321561; AAG50274.1; JOINED.  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.229; -  
 DR Genew: HGNC:6902; MASP2.  
 DR MIM: 605102; -  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000859; CUB\_domain.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR InterPro: IPR000436; Sush1\_SCR\_CCP.  
 DR Pfam: PF00084; sush1; 2.  
 DR Pfam: PF00089; trypsin; 1.  
 DR Pfam: PF00431; CUB; 2.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00032; CCP; 2.  
 DR SMART: SM00042; CUB; 2.  
 DR SMART: SM00179; EGF\_Ca; 1.  
 DR SMART: SM00020; TRYP\_SPC; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_Ca; 1.  
 DR PROSITE: PS02040; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_SER; FALSE\_NEG.  
 DR PROSITE: PS00135; TRYPSIN\_SRS; 1.  
 KW Hydroxylase, Complement pathway; Serine protease; Protease;  
 KW Glycoprotein; Sush1; Repeat; Signal; EGF-like domain; Hydroxylation.  
 KW SIGNAL  
 FT CHAIN 1 15  
 FT CHAIN 16 686  
 FT CHAIN 16 444  
 FT CHAIN 445 686  
 FT CHAIN 16 137  
 FT DOMAIN 138 181  
 FT DOMAIN 184 296  
 FT DOMAIN 299 362  
 FT DOMAIN 365 431  
 FT DOMAIN 445 686  
 FT ACT\_SITE 483 483  
 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 532 532 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 633 633 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT MOD\_RES 158 158 HYDROXYLATION (POTENTIAL).  
 FT DISULFID 142 156 POTENTIAL.  
 FT DISULFID 142 156 POTENTIAL.  
 FT DISULFID 152 165 POTENTIAL.  
 FT DISULFID 167 180 POTENTIAL.  
 FT DISULFID 184 211 POTENTIAL.  
 FT DISULFID 241 259 POTENTIAL.  
 FT DISULFID 300 348 POTENTIAL.  
 FT DISULFID 328 361 POTENTIAL.  
 FT DISULFID 366 412 POTENTIAL.  
 FT DISULFID 396 430 POTENTIAL.  
 FT DISULFID 434 552 INTERCHAIN (POTENTIAL).  
 FT DISULFID 598 618 POTENTIAL.  
 FT DISULFID 629 660 POTENTIAL.  
 FT CONFLICT 361 362 MISSING (IN REF. 3).  
 FT CONFLICT 371 371 D -> Y (IN REF. 4).  
 FT CONFLICT 372 372 L -> LCS (IN REF. 3).  
 FT CONFLICT 442 442 G -> E (IN REF. 4).  
 FT CONFLICT 447 447 G -> E (IN REF. 4).  
 FT CONFLICT 461 462 MISSING (IN REF. 3).  
 FT CONFLICT 473 473 L -> LIL (IN REF. 3).  
 SQ SEQUENCE 686 AA; 75685 MW; 4E34DED159448A2A CRC64;  
 Query Match 100.0%; Score 238; DB 1; Length 686;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-20;  
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TPLCPKWPPEVFGRLASPGPGGEYANDQERRWTLTAPPGYR 41  
 16 TPLGPKWPEEVEFGRLASPGPGGEYANDQERRWTLTAPPGYR 56  
 Db 16 TPLGPKWPEEVEFGRLASPGPGGEYANDQERRWTLTAPPGYR 56  
 RESULT 2  
 ID CRAR\_HUMAN STANDARD; PRT; 699 AA.  
 AC P48740; O95570; O9UFE09;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Complement-activating component of Ra-reactive factor precursor  
 DE (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (RARF)  
 DE (Mannan-binding lectin serine protease 1) (Mannose-binding protein  
 DE associated serine protease) (MASP-1).  
 GN MASP1 OR CRARF OR CRARF1 OR PRSS5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RC MEDLINE=94059062; PubMed=8240317;  
 RA Takada F., Takayama Y., Hattuse H., Kawakami M.;  
 RA "A new member of the Cls family of complement proteins found in a  
 RA bactericidal factor, Ra-reactive factor, in human serum.";  
 RL Biochem. Biophys. Res. Commun. 196:1003-1009(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal liver;  
 RC MEDLINE=94289349; PubMed=8018603;  
 RA Sato T., Endo Y., Matsushita M., Fujita T.;  
 RA "Molecular characterization of a novel serine protease involved in  
 RA activation of the complement system by mannose-binding protein.";  
 RL Int. Immunol. 6:665-669(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RC MEDLINE=97079701; PubMed=8921412;  
 RA Endo Y., Sato T., Matsushita M., Fujita T.;  
 RA "Exon structure of the gene encoding the human mannose-binding  
 RT protein-associated serine protease light chain: comparison with

RT complement C1r and C1s genes";  
RN Int. Immunol. 8:1355-1358(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99402590; PubMed-10475605;  
RA Takayama T., Takada F., Nowatari M., Kawakami M., Matsu-ura N.;  
RT "Gene structure of the p100 serine-protease component of the human Ra-  
reactive factor";  
RL Mol. Immunol. 36:505-514(1999).  
CC -1- FUNCTION: COMPONENT OF THE BACTERICIDIAL RA-REACTIVE FACTOR RARF  
WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY  
CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT  
CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE  
C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.  
CC -1- SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT  
(CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.  
CC CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (P29)  
LINKED BY A DISULFIDE BOND.  
CC -1- DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D17525; BAA04477.1; -  
DR EMBL: D28593; BAA05928.1; -  
DR EMBL: D61695; BAA34864.1; -  
DR EMBL: AB010822; BAA34864.1; JOINED.  
DR EMBL: AB010813; BAA34864.1; JOINED.  
DR EMBL: AB010814; BAA34864.1; JOINED.  
DR EMBL: AB010815; BAA34864.1; JOINED.  
DR EMBL: AB010816; BAA34864.1; JOINED.  
DR EMBL: AB010817; BAA34864.1; JOINED.  
DR EMBL: AB010818; BAA34864.1; JOINED.  
DR EMBL: AB010819; BAA34864.1; JOINED.  
DR EMBL: AB010820; BAA34864.1; JOINED.  
DR EMBL: AB010821; BAA34864.1; JOINED.  
DR EMBL: D61690; BAA34864.1; JOINED.  
DR EMBL: D61691; BAA34864.1; JOINED.  
DR EMBL: D61692; BAA34864.1; JOINED.  
DR EMBL: D61693; BAA34864.1; JOINED.  
DR EMBL: D61694; BAA34864.1; JOINED.  
DR EMBL: AB007617; BAA89206.1; -  
DR EMBL: AB007602; BAA89206.1; JOINED.  
DR EMBL: AB007603; BAA89206.1; JOINED.  
DR EMBL: AB007604; BAA89206.1; JOINED.  
DR EMBL: AB007605; BAA89206.1; JOINED.  
DR EMBL: AB007606; BAA89206.1; JOINED.  
DR EMBL: AB007607; BAA89206.1; JOINED.  
DR EMBL: AB007608; BAA89206.1; JOINED.  
DR EMBL: AB007609; BAA89206.1; JOINED.  
DR EMBL: AB007610; BAA89206.1; JOINED.  
DR EMBL: AB007611; BAA89206.1; JOINED.  
DR EMBL: AB007612; BAA89206.1; JOINED.  
DR EMBL: AB007613; BAA89206.1; JOINED.  
DR EMBL: AB007614; BAA89206.1; JOINED.  
DR EMBL: AB007615; BAA89206.1; JOINED.  
DR EMBL: AB007616; BAA89206.1; JOINED.  
DR HSSP: P00736; IAPQ.  
DR MEMOPS: S01.198; -  
DR Genew: HGNC:6901; MASPL.  
DR MIM: 600521; -  
DR InterPro: IPR000152; ASX\_hydroxyl.  
DR InterPro: IPR000859; CUB\_domain.  
DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR InterPro: IPR000436; Sush1\_SCR\_CCP.  
DR Pfam: PF00084; sush1; 2.  
DR Pfam: PF00089; trypsin; 1.  
DR Pfam: PF00431; CUB; 2.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00032; CCP; 2.  
DR SMART: SM00042; CUB; 2.  
DR SMART: SM00181; EGF; 1.  
DR SMART: SM00179; EGF\_CA; 1.  
DR SMART: SM00020; TRYP\_SPE; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE: PS01180; CUB; 2.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS01187; EGF\_CA; 1.  
DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
KW Hydrolyase: Complement pathway; Serine protease: Protease:  
KW Glycoprotein: Sush1; Repeat; Signal: EGF-like domain; Hydroxylation.  
FT SIGNAL 1 19  
FT CHAIN 20 699  
FT FT  
FT CHAIN 20 448  
FT CHAIN 449 699  
FT DOMAIN 20 138  
FT DOMAIN 139 182  
FT DOMAIN 185 297  
FT DOMAIN 300 363  
FT DOMAIN 366 433  
FT DOMAIN 449 699  
FT ACT\_SITE 490 490  
FT ACT\_SITE 532 532  
FT ACT\_SITE 646 646  
FT MOD\_RES 159 159  
FT DISULFID 73 91  
FT DISULFID 143 157  
FT DISULFID 153 166  
FT DISULFID 168 181  
FT DISULFID 185 212  
FT DISULFID 242 260  
FT DISULFID 301 349  
FT DISULFID 329 362  
FT DISULFID 367 414  
FT DISULFID 397 432  
FT DISULFID 436 572  
FT DISULFID 614 631  
FT DISULFID 642 672  
FT CARBOHYD 49 49  
FT CARBOHYD 178 178  
FT CARBOHYD 385 385  
FT CARBOHYD 407 407  
FT CONFLICT 235 235  
FT CONFLICT 285 285  
FT CONFLICT 499 499  
FT CONFLICT 499 499  
FT CONFLICT 527 527  
FT CONFLICT 543 543  
FT CONFLICT 552 552  
FT CONFLICT 643 643  
SQ SEQUENCE 699 AA: 79258 MW: 79258 MW: ADD9697AE6AB01B5 CRC64:  
Query Match 39.1%; Score 93; DB 1; Length 699;  
Best Local Similarity 45.2%; Pred. No. 0.0016;  
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;  
OY 11 VEGRLASPGFGEYANQDERWTLTAPGVR 41  
DB 27 MFGQISPGYPSDYSPDSEVTWNTVPDGR 57









CC -1- PTM: C-TERMINALLY PROCESSED AT MULTIPLE POSITIONS.  
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 NTR DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: L33799; AAA61949.1; ALT\_SEQ.  
CC EMBL: AB008549; BAA23281.1; -  
CC EMBL: AF053556; AAC78800.1; -  
CC EMBL: AF083655; AAD16041.1; -  
CC Genes: HGNC:8738; PCOLCE.  
CC MIM: 600270; -  
CC InterPro: IPR000859; CUB\_domain.  
CC InterPro: IPR001134; Netrin\_C.  
CC Pfam: PF00431; CUB; 2.  
CC Pfam: PF01759; NTR; 1.  
CC SMART: SM00042; CUB; 2.  
CC PROSITE: PS01180; CUB; 2.  
CC Glycoprotein: Repeat; Signal.  
CC SIGNAL 1 25 POTENTIAL.  
CC CHAIN 26 449 PROCOLLAGEN C-PROTEINASE ENHANCER  
CC FT 1 25 PROTEIN.  
CC FT DOMAIN 37 149 CUB 1.  
CC FT 159 273 CUB 2.  
CC FT DOMAIN 329 438 NTR.  
CC FT SITE 287 288 CLEAVAGE.  
CC FT SITE 288 289 CLEAVAGE.  
CC FT SITE 293 294 CLEAVAGE.  
CC FT SITE 299 300 CLEAVAGE.  
CC SITE 303 304 CLEAVAGE.  
CC CARBOHYD 29 304 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SO SEQUENCE 449 AA; 47972 MW; 3D88430158648796 CRC64;

Query Match 32.8%; Score 78; DB 1; Length 449;  
Best Local Similarity 37.8%; Pred. No. 0.056;  
Matches 17; Conservative 5; Mismatches 13; Indels 10; Gaps 1;

QY 5 PKRPEPVF-----GRLSPGFGEXANQERRWTLTAPP 39  
28 PNTRYFVLCGGDYKSGEGIVASEGPNLYPPKKECIWITVPEG 72

CIS\_HUMAN  
ID CIS\_HUMAN STANDARD: PRT: 688 AA.  
AC P09871; Q9UM14; Q9UCU7; Q9UCU8; Q9UCU9; Q9UCV0; Q9UCV1; Q9UCV2;  
AC Q9UCV3; Q9UCV4; Q9UCV5;  
AC 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Complement C1s component precursor (EC 3.4.21.42) (C1 esterase).  
GN C1S.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89017187; PubMed=2459702;  
RT Kusumoto H., Hirose S., Salier J.P., Hagen F.S., Kurauchi K.;  
RT "Human genes for complement component C1r and C1s in a close  
RT tail-to-tail arrangement.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:7307-7311(1988).  
RN [12]  
RP SEQUENCE FROM N.A.  
RA TISSUE=Liver;

RX MEDLINE=88082788; PubMed=3500856;  
RA McKinnon C.M., Carter P.E., Smyth S.J., Dunbar B., Fothergill J.E.;  
RT "Molecular cloning of cDNA for human complement component C1s. The  
RT complete amino acid sequence.";  
RL Eur. J. Biochem. 169:547-553(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88163522; PubMed=2831944;  
RA Tosti M., Duponchel C., Meo T., Julier C.;  
RT "Complete cDNA sequence of human complement C1s and close physical  
RT linkage of the homologous genes C1s and C1r.";  
RL Biochemistry 26:8516-8524(1987).  
RN [4]  
RP SEQUENCE OF 1-329 FROM N.A.  
RC TISSUE=Peripheral blood leukocytes;  
RX MEDLINE=99008558; PubMed=9794427;  
RA Endo Y., Takahashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,  
RA Nonaka M., Fujita T.;  
RT "Two lineages of mannose-binding lectin-associated serine protease  
RT (MASP) in vertebrates.";  
RL J. Immunol. 161:4924-4930(1998).  
RN [5]  
RP SEQUENCE OF 291-688 FROM N.A.  
RX MEDLINE=90040704; PubMed=2553984;  
RA Tosi M., Duponchel C., Meo T., Couture-Tosi E.;  
RT "Complement genes C1r and C1s feature an intronless serine protease  
RT domain closely related to haptoglobin.";  
RL J. Mol. Biol. 208:709-714(1989).  
RN [6]  
RP SEQUENCE OF 16-61; 168-219; 287-334 AND 384-445.  
RX MEDLINE=86164350; PubMed=3007145;  
RA Spycher S.E., Nick H., Rickli E.E.;  
RT "Human complement component C1s. Partial sequence determination of  
RT the heavy chain and identification of the peptide bond cleaved during  
RT activation.";  
RL Eur. J. Biochem. 156:49-57(1986).  
RN [7]  
RP SEQUENCE OF 438-500; 503-534; 542-601; 617-623 AND 626-656.  
RX MEDLINE=84104122; PubMed=6362661;  
RA Carter P.E., Dunbar B., Fothergill J.E.;  
RT "The serine proteinase chain of human complement component C1s.  
RT Cyanogen bromide cleavage and N-terminal sequences of the  
RT fragments.";  
RL Biochem. J. 215:565-571(1983).  
RN [8]  
RP PARTIAL SEQUENCE.  
RC TISSUE=Plasma;  
RX MEDLINE=91308095; PubMed=1854725;  
RA Ily C., Thielens N.M., Gagnon J., Arlaud G.J.;  
RT "Effect of lactoperoxidase-catalyzed iodination on the C4(2+)-  
RT dependent interactions of human C1s. Location of the iodination  
RT sites.";  
RL Biochemistry 30:7135-7141(1991).  
RN [9]  
RP DISULFIDE BONDS.  
RX MEDLINE=91175725; PubMed=2007122;  
RA Hess D., Schaller J., Rickli E.E.;  
RT "Identification of the disulfide bonds of human complement C1s.";  
RL Biochemistry 30:2827-2833(1991).  
RN [10]  
RP PARTIAL SEQUENCE, AND 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.  
RX MEDLINE=95288736; PubMed=779774;  
RA Rossi V., Gabrion C., Lacroix M., Ulrich J., Fontecilla-Camps J.C.,  
RA Gagnon J., Arlaud G.J.;  
RT "Structure of the catalytic region of human complement protease C1s:  
RT study by chemical cross-linking and three-dimensional homology  
RT modeling.";  
RL Biochemistry 34:7311-7321(1995).  
RN [11]  
RP DISEASE.  
RX PubMed=11390518;  
RA Dragon-Durey M.-A., Quartier P., Fremiaux-Bacchi V., Blouin J.,  
RA de Barce C., Prieur A.-M., Weiss L., Fridman W.-H.;

RT "Molecular basis of a selective C1s deficiency associated with early  
 RT onset multiple autoimmune diseases." ;  
 RL J. Immunol. 166:7612-7616(2001).  
 CC -1- FUNCTION: C1S B CHAIN IS A SERINE PROTEASE THAT COMBINES WITH C1O  
 CC AND C1S TO FORM C1, THE FIRST COMPONENT OF THE CLASSICAL PATHWAY  
 CC OF THE COMPLEMENT SYSTEM. C1R ACTIVATES C1S SO THAT IT CAN, IN  
 CC TURN, ACTIVATE C2 AND C4.  
 CC -1- CATALYTIC ACTIVITY: CLEAVES COMPONENT C4 TO C4A AND C4B, AND  
 CC COMPONENT C2 TO C2A AND C2B.  
 CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1O, R  
 CC AND S IN THE MOLAR RATION OF 1:2:2. ACTIVATED C1S IS AN DISULFIDE-  
 CC LINKED HETERODIMER OF AN HEAVY CHAIN AND A LIGHT CHAIN.  
 CC -1- DISEASE: Defects in C1S are the cause of selective C1s deficiency,  
 CC that is associated with early onset multiple autoimmune diseases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X06596; CAA29817.1; -  
 CC EMBL; J04080; AAA51852.1; -  
 CC EMBL; M18767; AAA51853.1; -  
 CC EMBL; AB09076; BAA86864.1; -  
 CC PIR; A40496; C1HUS.  
 CC PIR; S00224; S00224.  
 CC MEROPS: S01.193; -  
 CC SWISS-2PAGE; P09871; HUMAN.  
 CC Gene; HGNC:1247; C1S.  
 CC MIM; 120580; -  
 CC InterPro: IPR000152; Asx\_hydroxyl.  
 CC InterPro: IPR000859; CUB\_domain.  
 CC InterPro: IPR001314; Chymotrypsin.  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR001881; EGF\_Ca.  
 CC InterPro: IPR001254; Ser\_protease\_Try.  
 CC InterPro: IPR000436; Sushi\_SCR\_CCP.  
 CC Pfam; PF00084; EGF; 1.  
 CC Pfam; PF00089; sushi; 2.  
 CC Pfam; PF00089; trypsin; 1.  
 CC Pfam; PF00431; CUB; 2.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00032; CCP; 2.  
 CC SMART; SM00042; CUB; 2.  
 CC SMART; SM00179; EGF\_CA; 1.  
 CC SMART; SM00020; TRYP\_SPE; 1.  
 CC PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 CC PROSITE; PS01180; CUB; 2.  
 CC PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 CC PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
 CC PROSITE; PS01187; EGF\_CA; 1.  
 CC PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HTS; FALSE\_NEG.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Complement pathway; Plasma; Glycoprotein; Serine protease; Hydrolase;  
 CC Hydroxylation; Sushi; Repeat; Signal; EGF-like domain;  
 CC Calcium-binding.  
 CC SIGNAL 1 15  
 CC CHAIN 16 437 COMPLEMENT C1S HEAVY CHAIN.  
 CC CHAIN 438 688 COMPLEMENT C1S LIGHT CHAIN.  
 CC CHAIN 16 130 CUB 1.  
 CC DOMAIN 131 172 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).  
 CC DOMAIN 175 290 CUB 2.  
 CC DOMAIN 293 355 SUSHI 1.  
 CC DOMAIN 358 422 SUSHI 2.  
 CC DOMAIN 438 688 SERINE PROTEASE.

FT ACT\_SITE 475 475 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 529 529 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 632 632 CHARGE RELAY SYSTEM.  
 FT MOD\_RES 149 149 HYDROXYLATION (PROBABLE).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .).  
 FT DISULFID 65 83  
 FT DISULFID 135 147  
 FT DISULFID 143 156  
 FT DISULFID 158 171  
 FT DISULFID 175 202  
 FT DISULFID 234 251  
 FT DISULFID 294 341  
 FT DISULFID 321 354  
 FT DISULFID 359 403  
 FT DISULFID 366 421  
 FT DISULFID 425 549  
 FT DISULFID 595 618  
 FT DISULFID 628 659  
 FT DISULFID 294 294 C -> K (IN REF. 6).  
 FT CONFLICT 513 513 G -> GG (IN REF. 5).  
 FT CONFLICT 573 573 T -> A (IN REF. 7).  
 FT CONFLICT 645 646 TK -> GR (IN REF. 7).  
 SQ SEQUENCE 688 AA; 76684 MW; 85522647A4C47205 CRC64;  
 -----  
 QY 7 WPEP-VFGRLASPFGEVANDERMTLTAPPGY 40  
 Db 14 YAEPTWYGETLSPNYQPAVPSEVKSMDEVPFGY 48  
 -----  
 RESULT 8  
 ID PCOL\_MOUSE STANDARD; PRT; 468 AA.  
 AC 061398; 035113;  
 DT 30-MAY-2000 (Rel. 39, created)  
 DT 30-MAY-2000 (Rel. 39, last sequence update)  
 DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Procollagen C-proteinase enhancer protein precursor (PCPE) (Type I  
 DE procollagen COOH-terminal proteinase enhancer) (Type I procollagen C-  
 DE proteinase enhancer protein) (P14).  
 GN PCOLCE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=95014462; PubMed=7523404;  
 RA Takahara K., Kessler E., Biniatnov L., Brusel M., Eddy R.L.,  
 RA Jant-Salt S., Shows T.B., Greenspan D.S.;  
 RT "Type I procollagen COOH-terminal proteinase enhancer protein:  
 RT identification, primary structure, and chromosomal localization of the  
 RT cognate human gene (PCOLCE).";  
 RL J. Biol. Chem. 269:26280-26285(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Heart;  
 RA Hirahara I., Syoutuda K., Harada K., Tomita M., Urakami K., Terai H.,  
 RA Morisaki N., Saito Y.;  
 RT "Smooth muscle cell derived procollagen C-protease enhancer protein.";  
 RL Cell Struct. Funct. 21:662-662(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=CNS;  
 RX MEDLINE=91225708; PubMed=2027019;  
 RA Iecan E., Zelenika D., Iaine M.C., Rhyner T., Pessac B.;  
 RT "Isolation of a novel cDNA corresponding to a transcript expressed in  
 RT the choroid plexus and leptomeninges.";  
 RL J. Neurochem. 56:2133-2138(1991).

```

RN [4]
RP CHARACTERIZATION.
RX MEDLINE-90092088; PubMed-2689170;
RA Kessler E., Adar R.;
RT "Type I procollagen C-proteinase from mouse fibroblasts. Purification
RL and demonstration of a 55-kDa enhancer glycoprotein.";
RN Eur. J. Biochem. 186:115-121(1989).
RN [5]
RP TISSUE SPECIFICITY.
RC STRAIN-ICR.
RX MEDLINE-91076903; PubMed-2256940;
RA Kessler E., Mould A.P., Humes D.J.S.;
RT "Procollagen type I C-proteinase enhancer is a naturally occurring
RL connective tissue glycoprotein.";
RN Biochem. Biophys. Res. Commun. 173:81-86(1990).
RN [6]
RP CARBOHYDRATE-LINKAGE SITE ASN-28.
RA Kessler E.;
RT "Unpublished observations (Jan-2000).";
RL "- FUNCTION: BINDS TO THE COOH-TERMINAL PROPEPTIDE OF TYPE I
CC PROCOLLAGEN AND ENHANCES PROCOLLAGEN C-PROTEINASE ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN INTERSTITIAL CONNECTIVE TISSUES
CC LIKE TENDONS, CALVARIA, SKIN AND AT A LOWER LEVEL IN HEART AND
CC SKELETAL MUSCLE.
CC -1- PTM: PROCESSED FROM A 55 KDA FORM TO 36 AND 34 KDA FORMS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 NTR DOMAIN.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
CC IN POSITION 391 AND 445.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB008548; BAA23280.1; -
CC DR EMBL: X57337; CAA0612.1; ALT_FRAME.
CC DR MGI: 105099; PcoIce.
CC DR InterPro: IPR000859; CUB_domain.
CC DR InterPro: IPR001134; Netrin_C.
CC DR Pfam: PF00431; CUB; 2.
CC DR Pfam: PF01759; NTR; 1.
CC DR SMART: SM00042; CUB; 2.
CC DR PROSITE: PS01180; CUB; 2.
CC DR PROSITE: PS01180; CUB; 2.
CC DR glycoprotein; Repeat; Signal.
CC SIGNAL 1 24
CC CHAIN 25 468
CC FT DOMAIN 36 148
CC FT DOMAIN 158 272
CC FT DOMAIN 352 461
CC FT CARBOHYD 28 28
CC FT CARBOHYD 454 454
CC FT CONFLICT 178 178
CC FT D -> Y (IN REF. 3) (POTENTIAL).
CC SO SEQUENCE 468 AA; 50168 MW; F9B5558147FAAB CRC64;

Query Match 31.1%; Score 74; DB 1; Length 468;
Best Local Similarity 35.6%; Pred. No. 0.17;
Matches 16; Conservative 6; Mismatches 13; Indels 10; Gaps 1;

QY 5 PKPPEPVF-----GRLASPGFPGGEYANDQERRWITLAPPG 39
ID 1 111 1 111 111 1 111 111 1
Db 27 PNTYRPVFLGGDVTGESGVASEGPNLYPKKKCIWITIVEG 71

RESULT 9
PCOL_RAT STANDARD: PRT; 468 AA.
AC 008628;

```

```

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Procollagen C-proteinase enhancer protein precursor (PCPE) (Type I
DE procollagen COOH-terminal proteinase enhancer) (Type I procollagen C-
DE proteinase enhancer protein).
GN PCOLCE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Mistar;
RX MEDLINE-97447750; PubMed-9303490;
RA Ogata I., Auster A.S., Matsui A., Greenwel P., Ceerts A., D'Amico T.,
RA Fujiwara K., Kessler E., Rojkind M.;
RT "Up-regulation of type I procollagen C-proteinase enhancer protein
RL messenger RNA in rats with CCl4-induced liver fibrosis.";
RN Hepatology 26:611-617(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Mistar; TISSUE-Aorta;
RA Hirahara I., Syoutuda K., Harada K., Tomita M., Urakami K., Terai H.,
RA Morisaki N., Saito Y.;
RT "Smooth muscle cell derived procollagen C-protease enhancer protein.";
RL Cell Struct. Funct. 21:662-662(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Fischer 344;
RA Masuda M., Igarashi H., Kano M., Yoshikura H.;
RT "Proviral integration into the procollagen C-proteinase enhancer
RL protein gene and its effects in cultured rat fibroblasts revealed by
RL an excisable 'hit-and-run' retroviral vector.";
RN Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP TISSUE SPECIFICITY.
RA Kessler E., Mould A.P., Humes D.J.S.;
RX MEDLINE-91076903; PubMed-2256940;
RT "Procollagen type I C-proteinase enhancer is a naturally occurring
RL connective tissue glycoprotein.";
RN Biochem. Biophys. Res. Commun. 173:81-86(1990).
CC -1- FUNCTION: BINDS TO THE COOH-TERMINAL PROPEPTIDE OF TYPE I
CC PROCOLLAGEN AND ENHANCES PROCOLLAGEN C-PROTEINASE ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN COLLAGEN-RICH
CC TISSUES, ESPECIALLY TENDON. ALSO EXPRESSED IN CORNEA AND STERNA.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 NTR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U94710; AAB93478.1; -
CC DR EMBL: AB008534; BAA23217.1; -
CC DR EMBL: AF016503; AAD01592.1; -
CC DR EMBL: AF016506; AAD01598.1; -
CC DR InterPro: IPR000859; CUB_domain.
CC DR InterPro: IPR001134; Netrin_C.
CC DR Pfam: PF00431; CUB; 2.
CC DR Pfam: PF01759; NTR; 1.
CC DR SMART: SM00042; CUB; 2.
CC DR PROSITE: PS01180; CUB; 2.
CC DR glycoprotein; Repeat; Signal.
CC FT SIGNAL 1 24
CC FT CHAIN 25 468
CC FT DOMAIN 36 148
CC FT CUB 1.
CC PROCOLLAGEN C-PROTEINASE ENHANCER
CC PROTEIN.

```

FT DOMAIN 158 272 CUB 2.  
FT DOMAIN 352 461 NTR.  
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 468 AA; 50185 MW; B4AA1C151323969B CRC64;  
Query Match 31.1%; Score 74; DB 1; Length 468;  
Best Local Similarity 35.6%; Pred. No. 0.17;  
Matches 16; Conservative 6; Mismatches 13; Indels 10; Gaps 1;  
QY 5 PKMPEPYF-----GLASPGFGEVANDQERMTITAPG 39  
DB 27 PNTRYFVLCGGDVTEGSGVASEGPNLYPPKKCIWITVPEG 71  
RESULT 10  
UWS2\_XENLA STANDARD; PRT; 514 AA.  
AC P42664;  
01-NOV-1995 (Rel. 32, Created)  
11-NOV-1997 (Rel. 35, Last sequence update)  
05-JUN-2002 (Rel. 41, Last annotation update)  
Embryonic protein UWS.2 precursor (EC 3.4.24.-).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_Taxid=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=97228908; PubMed=9074934;  
RA Katsuragi C., Maeda R., Yamashika C., Mita K., Sargent T.D.,  
RA Yasumasa S.;  
RT "Molecular cloning of Xenopus hatching enzyme and its specific  
RT expression in hatching gland cells.";  
RL Int. J. Dev. Biol. 41:13-25(1997).  
RN [2]  
RP SEQUENCE OF 196-514 FROM N.A.  
RA MEDLINE=90108413; PubMed=1688538;  
RA Sato S.M., Sargent T.D.;  
RT "Molecular approach to dorsoanterior development in Xenopus laevis.";  
RL Dev. Biol. 137:135-141(1990).  
RN [3]  
RP COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -1- DEVELOPMENTAL STAGE: EXCLUSIVELY IN THE ANTERIOR NEURAL FOLD OF  
CC NEURULA STAGE EMBRYOS. BY THE TAILBUD STAGE, THE PROTEIN IS  
CC LOCALIZED IN SPECIALIZED CEPHALIC ECTODERM, IN A REGION PROBABLY  
CC CORRESPONDING TO THE HATCHING GLAND.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.  
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: D89632; BAA14003.1; -  
DR EMBL: M27162; AAA49980.1; -  
DR HSSP: P07384; 1IAE.  
DR MEROPS: M12.014; -  
DR InterPro: IPR001506; Astacin.  
DR InterPro: IPR000859; CUB domain.  
DR InterPro: IPR000130; Zn\_MTPeptidase.  
DR Pfam: PF00431; CUB; 2.  
DR Pfam: PF01400; Astacin; 1.  
DR PRINTS: PR00480; ASTACIN.  
DR SMART: SM00042; CUB; 2.  
DR SMART: SM00235; ZMC; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
DR PROSITE: PS01180; CUB; 2.  
KW Hydrolase; Metalloprotease; Zinc; Repeat; Signal.

FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 514 EMBRYONIC PROTEIN UVS.2.  
FT DOMAIN 288 400 CUB 1.  
FT DOMAIN 402 513 CUB 2.  
FT METAL 186 186 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 187 187 BY SIMILARITY.  
FT METAL 190 190 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 196 196 ZINC (CATALYTIC) (BY SIMILARITY).  
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 514 AA; 56844 MW; DELFCB599014D986 CRC64;  
Query Match 31.1%; Score 74; DB 1; Length 514;  
Best Local Similarity 44.0%; Pred. No. 0.19;  
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 17 SPGPGFGEVANDQERMTITAPG 41  
DB 415 SPVPGVTTNTCTWITAPAGFK 439  
RESULT 11  
TLD\_BRARE STANDARD; PRT; 1022 AA.  
ID TLD\_BRARE  
AC 057460;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Dorsal-ventral patterning tolloid protein precursor (EC 3.4.24.-)  
DE (Mini fin protein).  
GN TOLLOID OR TLD OR MFN.  
OS Brachydanio rerio (zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_Taxid=7955;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
RC TISSUE=gastrula;  
RA MEDLINE=98057457; PubMed=9395394;  
RA Blader P., Rastegar S., Fischer N., Straehle U.;  
RT "Cleavage of the BMP-4 antagonist chordin by zebrafish Tolloid.";  
RL Science 278:1937-1940(1997).  
RN [2]  
RP FUNCTION, AND TISSUE SPECIFICITY.  
RA MEDLINE=99307076; PubMed=10375503;  
RA Connors S.A., Trout J., Ekker M., Mullins M.C.;  
RT "The role of tolloid/minf in dorsoventral pattern formation of the  
RT zebrafish embryo.";  
RL Development 126:3119-3130(1999).  
CC -1- FUNCTION: Required for patterning ventral tissues of the tail. May  
CC increase bone morphogenetic protein (BMP) activity at the end of  
CC gastrulation by proteolytic cleavage of chordin and release of BMP  
CC from inactive complexes.  
CC -1- TISSUE SPECIFICITY: During gastrulation, accumulates around the  
CC closing blastopore with greater expression ventrally. At the  
CC animal pole, expressed in the ectoderm flanking the anterior  
CC neural plate. At the 10-somite stage, expressed in the developing  
CC tailbud and cranial neural crest. At the 20-somite stage, also  
CC expressed in the hematopoietic system.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.  
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 5 CUB DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).

CC EMBL: AF027596; AAC60304.1; -

DR HSPSP: P35555; 1EMN.

DR MEROPS: M12.016; -

DR ZFIN: ZDB-GENE-990415-265; tollold.

DR InterPro: IPR001506; Asx\_hydroxyl.

DR InterPro: IPR000152; Asx\_hydroxyl.

DR InterPro: IPR000859; CUB-domain.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001881; EGF\_Ca.

DR InterPro: IPR000130; Zn\_Mpdpdase.

DR Pfam: PF00008; EGF; 2.

DR Pfam: PF00431; CUB; 5.

DR Pfam: PF01400; Asxacin; 1.

DR PRINTS: PR00480; ASTACIN.

DR SMART: SM00042; CUB; 5.

DR SMART: SM00181; EGF; 2.

DR SMART: SM00179; EGF\_Ca; 2.

DR SMART: SM00235; Zmnc; 1.

DR PROSITE: PS00010; ASX\_HYDROXYL; 2.

DR PROSITE: PS01180; CUB; 5.

DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.

DR PROSITE: PS01186; EGF\_2; 2.

DR PROSITE: PS01187; EGF\_Ca; 2.

DR PROSITE: PS00142; ZINC\_PROTEASE; 1.

DR Developmental protein: Protease; Metalloprotease; Zinc;

KM Metal-binding; Calcium; EGF-like domain; Repeat; Signal; Glycoprotein;

KM Zymogen.

FT SIGNAL 1 32 POTENTIAL.

FT PROPEP 33 156 POTENTIAL.

FT CHAIN 157 1022 DORSAL-VENTRAL PATTERNING TOLLOID PROTEIN.

FT DOMAIN 157 357 METALLOPROTEASE (BY SIMILARITY).

FT DOMAIN 358 470 CUB 1.

FT DOMAIN 471 583 CUB 2.

FT DOMAIN 583 624 EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 627 739 CUB 3.

FT DOMAIN 739 779 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 783 895 CUB 4.

FT DOMAIN 896 1012 CUB 5.

FT METAL 249 249 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT\_SITE 250 250 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 253 253 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 259 259 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 358 384 BY SIMILARITY.

FT METAL 411 433 BY SIMILARITY.

FT METAL 497 546 BY SIMILARITY.

FT METAL 524 546 BY SIMILARITY.

FT METAL 567 599 BY SIMILARITY.

FT METAL 595 608 BY SIMILARITY.

FT METAL 610 623 BY SIMILARITY.

FT METAL 627 653 BY SIMILARITY.

FT METAL 702 702 BY SIMILARITY.

FT METAL 743 754 BY SIMILARITY.

FT METAL 750 763 BY SIMILARITY.

FT METAL 765 778 BY SIMILARITY.

FT METAL 809 858 BY SIMILARITY.

FT METAL 836 896 BY SIMILARITY.

FT METAL 926 975 BY SIMILARITY.

FT METAL 953 975 BY SIMILARITY.

FT METAL 129 129 N-LINKED (GLCNAC; . . .) (POTENTIAL).

FT METAL 178 178 N-LINKED (GLCNAC; . . .) (POTENTIAL).

FT METAL 368 368 N-LINKED (GLCNAC; . . .) (POTENTIAL).

FT METAL 399 399 N-LINKED (GLCNAC; . . .) (POTENTIAL).

FT METAL 635 635 N-LINKED (GLCNAC; . . .) (POTENTIAL).

FT METAL 635 635 N-LINKED (GLCNAC; . . .) (POTENTIAL).

SO SEQUENCE 1022 AA; 115536 MW; A68CALDDE41793F9 CRC64;

Query Match 31.1%; Score 74; DB 1; Length 1022;

Best Local Similarity 41.4%; Pred. No. 0.38;

Matches 12; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

OY 13 GRASPPEYANDQERRMTLAPPGYR 41

DB 792 GTISSPMPDKIPKREKCTWIDITATPGHR 820

RESULT 12

CASP\_MESAU STANDARD; PRT; 695 AA.

AC P15156.

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Calcium-dependent serine proteinase precursor (EC 3.4.21.-) (CASP).

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

OC Mesocricetus.

OX NCBI\_TaxId=10036;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-50 AND 446-472.

RC TISSUE=Fibroblast; PubMed=2753140;

RX MEDLINE=89325606; PubMed=2753140;

RA Kinoshita H., Sakiyama H., Tokunaga K., Imajob-Ohmi S., Hamada Y.,

RA Isono K., Sakiyama S.;

RT "Complete primary structure of calcium-dependent serine proteinase capable of degrading extracellular matrix proteins.";

RL FEBS Lett. 250:411-415(1989).

CC -I- FUNCTION: CAPABLE OF DEGRADING EXTRACELLULAR MATRIX PROTEINS.

CC -I- CASP DEGRADABLE TYPE I AND IV COLLAGEN AND FIBRONECTIN IN THE PRESENCE OF CALCIUM.

CC -I- SUBUNIT: HETERODIMER, CONSISTING OF HEAVY AND LIGHT CHAINS WITH DISULFIDE BONDS. THE HEAVY CHAIN IS EXPECTED TO BE A REGULATORY SUBUNIT AND THE LIGHT CHAIN CONTAINS THE CATALYTIC SITE.

CC -I- DOMAIN: THE GLU-RICH REGION IN THE N-TERMINAL REGION MAY BE GAMMA CARBOXYLATED AND FUNCTION AS A CALCIUM-BINDING SITE.

CC -I- SIMILARITY: TO BLOOD COAGULATION FACTORS SUCH AS IX, X AND AN ANTICOAGULATION FACTOR, PROTEIN C.

CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC -I- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.

CC -I- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.lsb-sib.ch/announce/> or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

CC -----

CC EMBL: X16160; CA34286.1; -

CC PIR: S05008; S05008.

CC HSP: P00763; IDPO.

CC MEROPS: S01.193; -

CC InterPro: IPR000152; Asx\_hydroxyl.

CC InterPro: IPR000859; CUB-domain.

CC InterPro: IPR001314; Chymotrypsin.

CC InterPro: IPR000561; EGF-like.

CC InterPro: IPR001881; EGF\_Ca.

CC InterPro: IPR001254; Ser-protease\_Try.

CC InterPro: IPR000436; Sushi\_SCR\_CCP.

CC Pfam: PF00008; EGF; 1.

CC Pfam: PF00084; sushi; 2.

CC Pfam: PF00089; trypsin; 1.

CC Pfam: PF00431; CUB; 2.

CC PRINTS: PR00732; CHYMOTRYPSIN.

CC SMART: SM00033; CCP; 2.

CC SMART: SM00042; CUB; 2.

CC SMART: SM00179; EGF\_Ca; 1.

CC SMART: SM00020; Tryp\_Spc; 1.

CC PROSITE: PS00010; ASX\_HYDROXYL; 1.

CC PROSITE: PS01180; CUB; 2.

```

DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; FALSE_NEG.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Calcium-binding; Extracellular matrix;
KW Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation; Glycoprotein;
KW Signal; EGF-like domain; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 695 CALCIUM-DEPENDENT SERINE PROTEINASE.
FT CHAIN 22 444 CALCIUM-DEPENDENT SERINE PROTEINASE HEAVY
FT CHAIN 445 695 CHAIN.
FT CHAIN 445 695 CALCIUM-DEPENDENT SERINE PROTEINASELIGHT
FT DOMAIN 22 124 GLU-RICH (ACIDIC).
FT DOMAIN 22 136 CUB 1.
FT DOMAIN 181 296 CUB 2.
FT DOMAIN 237 349 SUSH1 1.
FT DOMAIN 352 415 SUSH1 2.
FT DOMAIN 445 695 SERINE PROTEASE.
FT DOMAIN 137 178 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
FT MOD_RES 155 155 HYDROXYLATION (POTENTIAL).
FT ACT_SITE 482 482 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 536 536 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 638 638 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 141 153 BY SIMILARITY.
FT DISULFID 149 162 BY SIMILARITY.
FT DISULFID 164 177 BY SIMILARITY.
FT CARBOHD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 695 AA; 77493 MW; E924F7E6340700D0 CRC64;

Query Match 30.5%; Score 72.5; DB 1; Length 695;
Best Local Similarity 36.4%; Pred No. 0.39;
Matches 12; Conservative 7; Mismatches 13; Indels 1; Gaps 1;

OY 9 EP-VFGRLASPGFPGPEYANDQERMTLAPGY 40
DB 22 EPTMGEILSPNVPQAYPMEMKTDIEVEGFG 54

RESULT 13
BMP1_HUMAN STANDARD; PRT; 986 AA.
AC P13497; Q13492; Q99421; Q99422; Q14874; Q9UL38; Q13872;
DE 01-JAN-1990 (Rel. 13, Created)
PE 16-OCT-2001 (Rel. 40, Last sequence update)
PF 15-JUN-2002 (Rel. 41, Last annotation update)
PR Bone morphogenetic protein 1 precursor (EC 3.4.24.19) (BMP-1)
PT Procollagen C-proteinase (PCP) (Mammalian tollloid protein) (mtld).
GN BMP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM BMP1-3).
RC TISSUE=Skin;
RX MEDLINE=96209868; PubMed=8643539;
RA Li S.W., Steron A.L., Fertala A., Hojima Y., Arnold W.V.,
RA Prockop D.J.;
RA "The C-proteinase that processes procollagens to fibrillar collagens
RT is identical to the protein previously identified as bone morphogenic
RL protein-1."
RL Proc. Natl. Acad. Sci. U.S.A. 93:5127-5130(1996).
RN [2]
RN SEQUENCE FROM N.A. (ISOFORM BMP1-1).
RX MEDLINE=89072730; PubMed=3201241;
RA Mooney J.M., Rosen V., Celeste A.J., Mitscock L.M., Whitters M.J.,
RA Kritz R.W., Hewick R.M., Wang E.A.;
RA "Novel regulators of bone formation: molecular clones and
RT activities."

```

```

RL Science 242:1528-1534(1988).
RN [3]
RN SEQUENCE FROM N.A. (ISOFORMS BMP1-4; BMP1-5 AND BMP1-6).
RP TISSUE=Placenta;
RC MEDLINE=98160316; PubMed=9500680;
RX Janitz M., Helser V., Boeltcher U., Landt O., Lauster R.;
RA "Three alternatively spliced variants of the gene coding for the human
RT bone morphogenetic protein-1."
RL J. Mol. Med. 76:141-146(1998).
RN [4]
RN PARTIAL SEQUENCE FROM N.A. (ISOFORMS BMP1-3 AND BMP1-7).
RP TISSUE=Placenta;
RX MEDLINE=95096114; PubMed=7798260;
RA Takahara K., Lyons G.E., Greenspan D.S.;
RT "Bone morphogenetic protein-1 and a mammalian tollloid homologue (mtld)
RT are encoded by alternatively spliced transcripts which are
RT differentially expressed in some tissues."
RL J. Biol. Chem. 269:32572-32578(1994).
CC -1- FUNCTION: CLEAVES THE C-TERMINAL PROPEPTIDES OF PROCOLLAGEN I, II
CC AND II. INDUCES CARTILAGE AND BONE FORMATION.
CC -1- CATALYTIC ACTIVITY: Cleavage of the C-terminal propeptide at
CC Ala-I-Asp in type I and II procollagens and at Arg-I-Asp in type
CC III.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- ENZYME REGULATION: ACTIVITY IS INCREASED BY THE PROCOLLAGEN C-
CC ENDOPEPTIDASE ENHANCER PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 7 isoforms; BMP1-1, BMP1-2, BMP1-3 (shown
CC here); BMP1-4, BMP1-5, BMP1-6 and BMP1-7; are produced by
CC alternative splicing.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 CUB DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U50330; AAA93462.1; -
DR EMBL: M22488; AAA51833.1; -
DR EMBL: Y08723; CAA69974.1; -
DR EMBL: Y08724; CAA69974.1; -
DR EMBL: Y08725; CAA69975.1; -
DR EMBL: L35278; AAC41703.1; -
DR EMBL: L35279; AAC41710.1; -
DR PIR: A37278; A37278.
DR HSSP: P00736; IAPQ.
DR MEROPS: M12.005; -.
DR Genew: HGNC:1067; BMP1.
DR MIM: 112264; -.
DR InterPro: IPR001506; Astacin.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF_1like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR000130; Zn_MTPeptidse.
DR Pfam: PF00008; EGF_2.
DR Pfam: PF00431; Astc; 5.
DR Pfam: PF01400; Astacin; 1.
DR PRINTS: PR00480; ASTACIN.
DR SMART: SM00042; CUB; 5.
DR SMART: SM00179; EGF_CA; 2.
DR SMART: SM00235; ZnMC; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS01180; Asx; 5.
DR PROSITE: PS00010; ASX_HYDROXYL; 2.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01187; EGF_CA; 2.

```



KW Growth factor; Cytokine; Repeat; Bone; Cartilage; Hydrolase;  
 KM Metalloprotease; EGF-like domain; Zinc; Calcium; Signal;  
 FT Glycoprotein; Zymogen; Alternative splicing.  
 FT SIGNAL 1 22  
 FT PROPEP 1 120  
 FT CHAIN 121 966  
 FT DOMAIN 121 966  
 FT DOMAIN 322 434  
 FT DOMAIN 435 546  
 FT DOMAIN 547 588  
 FT DOMAIN 591 703  
 FT DOMAIN 704 743  
 FT DOMAIN 747 759  
 FT DOMAIN 859 976  
 FT METAL 213 213  
 FT ACT\_SITE 214 217  
 FT METAL 217 217  
 FT METAL 223 223  
 FT METAL 223 223  
 FT DISULFID 332 348  
 FT DISULFID 375 397  
 FT DISULFID 435 461  
 FT DISULFID 488 510  
 FT DISULFID 551 563  
 FT DISULFID 559 572  
 FT DISULFID 574 587  
 FT DISULFID 591 617  
 FT DISULFID 644 666  
 FT DISULFID 707 718  
 FT DISULFID 714 727  
 FT DISULFID 729 742  
 FT CARBOHYD 91 91  
 FT CARBOHYD 142 142  
 FT CARBOHYD 332 332  
 FT CARBOHYD 363 363  
 FT CARBOHYD 599 599  
 FT VARSPLIC 703 730  
 FT VARSPLIC 731 966  
 FT VARSPLIC 245 302  
 FT VARSPLIC 303 966  
 FT VARSPLIC 589 622  
 FT VARSPLIC 623 966  
 FT VARSPLIC 703 717  
 FT VARSPLIC 718 966  
 FT VARSPLIC 703 823  
 FT VARSPLIC 824 966  
 FT CONFLICT 748 748  
 FT CONFLICT 934 934  
 FT SEQUENCE 986 AA; 11248 MM; F89201913AC3CBEA CRC64;

Query Match 29.0%; Score 69; DB 1; Length 986;  
 Best Local Similarity 44.8%; Pred No. 14;  
 Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

RESULT 14  
 BMP1\_MOUSE

ID BMP1\_MOUSE STANDARD; PRT; 991 AA.  
 AC P98063;  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Bone morphogenetic protein 1 precursor (BC 3.4.24.19) (BMP-1)  
 DE (Procollagen C-proteinase) (PCP) (Mammalian tolloid protein) (mtld).  
 GN BMP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RX MEDLINE=94229342; PubMed=8174772;  
 RA Fukagawa M., Noboru S., Hogan B.L.M., Jones C.M.;  
 RT "Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1),  
 RT which is related to the Drosophila dorsoventral gene tolloid and  
 RT encodes a putative astacin metalloendopeptidase.";  
 RL Dev. Biol. 163:175-183(1994).  
 CC - FUNCTION: CLEAVES THE C-TERMINAL PROPEPTIDES OF PROCOLLAGEN I, II  
 CC AND II. INDUCES CARTILAGE AND BONE FORMATION.  
 CC - CATALYTIC ACTIVITY: Cleavage of the C-terminal propeptide at  
 CC Ala-I-Asp in type I and II procollagens and at Arg-I-Asp in type  
 CC III.  
 CC - COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC - ENZYME REGULATION: ACTIVITY IS INCREASED BY THE PROCOLLAGEN C-  
 CC ENDOPEPTIDASE ENHANCER PROTEIN.  
 CC - TISSUE SPECIFICITY: AT HIGH LEVELS IN EMBRYONIC MATERNAL DECIDUUM  
 CC AND FLOOR PLATE REGION OF THE NEURAL TUBE. LESS IN DEVELOPING  
 CC MEMBRANOUS AND ENDOCHONDRAL BONE. SUBMUCOSA OF INTESTINE, DERMIS  
 CC OF SKIN AND THE MESENCHYME OF SPLEEN AND LUNG.  
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1ZA.  
 CC - SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
 CC - SIMILARITY: CONTAINS 5 CUB DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L24755; AAA37306.1; -  
 DR HSSP; P00736; IAPQ.  
 DR MEROPS; M12.005; -  
 DR MGD; MGI:88176; Bmp1.  
 DR InterPro: IPR001506; Astacin.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000859; CUB\_domain.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR000130; Zn\_MTPeptidase.  
 DR Pfam: PR00008; EGF; 2.  
 DR Pfam: PR00431; CUB; 5.  
 DR Pfam: PR01400; Astacin; 1.  
 DR PRINTS; PR00480; ASTACIN.  
 DR SMART; SM00042; CUB; 5.  
 DR SMART; SM00179; EGF\_CA; 2.  
 DR SMART; SM00235; ZMNC; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE; PS01180; CUB; 5.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 2.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01187; EGF\_CA; 2.  
 KW Growth factor; Cytokine; Repeat; Bone; Cartilage; Hydrolase;  
 KW Metalloprotease; EGF-like domain; Zinc; Calcium; Signal;  
 KW Glycoprotein; Zymogen.  
 FT SIGNAL 1 25  
 FT PROPEP 26 125  
 FT POTENTIAL.  
 FT BMP1\_MOUSE



Query Match	Best Local Similarity	Score	DB	Length
Matches 11: Conservative	5: Mismatches	11: Indels	0: Gaps	0: Gaps

DB	316	GNFSSPEMPGQYDGTQTCVWRISTYTPG	342
RESULT	17		
NRPL_CHICK			
ID	NRPL_CHICK	STANDARD:	PRT: 914 AA.
AC	P79795.		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Neurophilin-1 precursor (A5 protein).		
GN	NRPL OR NRP.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OX	Gallus.		
NCBI_TaxID=9031;			
[1]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=White Leghorn; TISSUE=Embryonic brain;		
RX	MEDLINE=95324761; PubMed=7601310;		
RA	Takagi S., Kasuya Y., Shimizu M., Matsura T., Tsuboi M., Kawakami A., Fujisawa H.,		
RT	"Expression of a cell adhesion molecule, neurophilin, in the		
RL	developing chick nervous system";		
dev. Biol. 170:207-222(1995)."			
CC	- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE		
CC	CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF		
CC	CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS		
CC	SYSTEM. IT MODULATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY		
CC	SIMILARITY). SEEMS TO HAVE CALCIUM-INDEPENDENT CELL ADHESION		
CC	PROPERTIES.		
CC	- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	- TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM		
CC	(LAYERS D AND E OF SGF); AMACRINE CELLS OF RETINA, NEURITES OF		
CC	DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.		
CC	BLOOD VESSELS IN THE ENTIRE EMBRYO.		
CC	- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.		
CC	- SIMILARITY: CONTAINS 2 CUB DOMAINS.		
CC	- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.		
CC	- SIMILARITY: CONTAINS 1 MAM DOMAIN.		
CC	-----		
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL: D45416; BAA08356.1; .		
DR	HSSP: P12259; ICGT.		
DR	InterPro: IPR0000859; CUB_domain.		
DR	InterPro: IPR00421; FA5B_C.		
DR	InterPro: IPR000998; MAM_domain.		
DR	Pfam: PF00431; CUB; 2.		
DR	Pfam: PF00629; MAM; 1.		
DR	Pfam: PF00754; F5_F8_type_C; 2.		
DR	PRINTS: PR00020; MAMDOMAIN.		
DR	SMART: SM00042; CUB; 2.		
DR	SMART: SM00231; FA5B; 2.		
DR	SMART: SM00137; MAM; 1.		
DR	PROSITE: PS01180; CUB; 2.		
DR	PROSITE: PS01285; FA5B_C.1; 2.		
DR	PROSITE: PS01286; FA5B_C.2; 2.		
DR	PROSITE: PS00740; MAM_1; 1.		
DR	PROSITE: PS50060; MAM_2; 1.		
KW	Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;		
KW	Cell adhesion.		
FT	SIGNAL	1	18
FT	CHAIN	19	914
FT	DOMAIN	20	847
			POTENTIAL.
			NEUROFILIN-1.
			EXTRACELLULAR (POTENTIAL).

```

FT TRANSMEM 848 870 POTENTIAL.
FT DOMAIN 871 914 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 25 139 CUB 1.
FT DOMAIN 145 263 CUB 2.
FT DOMAIN 273 422 F5/8 TYPE C 1.
FT DOMAIN 429 581 F5/8 TYPE C 2.
FT DOMAIN 636 801 MAM.
FT DISULFID 25 52 PROBABLE.
FT DISULFID 80 102 PROBABLE.
FT DISULFID 145 171 PROBABLE.
FT DISULFID 204 226 PROBABLE.
FT DISULFID 273 422 BY SIMILARITY.
FT DISULFID 429 581 BY SIMILARITY.
SQ SEQUENCE 914 AA: 102480 MW: DD2E6D6FCBB68C CRC64:

Query Match 27.7% Score 66; DB 1; Length 914;
Best Local Similarity 41.4%; Pred. No. 2.9;
Matches 12; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

13 GRASPGEYANDQERRMTLTAPGGR 41
| | | | | | | | | | | | | | | |
35 GYLTPSPGQSYHPSQKCEWLQAPERYQ 63

RESULT 18
CA19_CHICK STANDARD; PRT; 503 AA.
ID CA19_CHICK
AC P12106;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE "collagen alpha 1(IX) chain precursor (Fragments)."
GN COL9A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP MEDLINE=90062114; PubMed=2584206;
RA Nishimura I., Muregaki Y., Olsen B.R.;
RT "Tissue-specific forms of type IX collagen-proteoglycan arise from
the use of two widely separated promoters."
J. Biol. Chem. 264:20033-20041(1989).
RN [2]
RP SEQUENCE OF 1-299 FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE=88115376; PubMed=3339014;
RA Sasaki G., Nishimura I., Konomi H., van der Rest M., Ninomiya Y.,
Olsen B.R.;
RT "Cartilage type IX collagen-proteoglycan contains a large
amino-terminal globular domain encoded by multiple exons."
J. Biol. Chem. 263:2324-2329(1988).
RN [3]
RP SEQUENCE OF 417-504 FROM N.A.
MEDLINE=85216631; PubMed=3858862;
RA Lozano G., Ninomiya Y., Thompson H., Olsen B.R.;
RT "A distinct class of vertebrate collagen genes encodes chicken type
IX collagen polypeptides."
Proc. Natl. Acad. Sci. U.S.A. 82:4050-4054(1985).
RN [4]
RP FUNCTION: Structural component of hyaline cartilage and vitreous
of the eye.
CC -1- SUBUNIT: Heterotrimer of a alpha 1(IX), a alpha 2(IX) and a alpha
3(IX) chain.
CC -1- DOMAIN: EACH SUBUNIT IS COMPOSED OF THREE TRIPLE-HELICAL DOMAINS
INTERSPERSED WITH NONCOLLAGENOUS DOMAINS. THE GLOBULAR DOMAIN AT
THE N-TERMINUS OF TYPE IX COLLAGEN MOLECULES REPRESENTS THE NC4
DOMAIN WHICH MAY PARTICIPATE IN ELECTROSTATIC INTERACTIONS WITH
POLYANIONIC GLYCOSAMINOGLYCANS IN CARTILAGE.
CC -1- PTM: Covalently linked to the telopeptides of type II collagen by
lysine-derived cross-links.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

```

```

CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
INTERPRETED HELICES (FACIT) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M28659; AAA48708.1; -
DR EMBL; J03539; AAA48702.1; -
DR EMBL; M11325; AAA48642.1; -
DR EMBL; M11324; AAA48642.1; JOINED.
DR PIR; A28360; A28360.
DR PIR; A28754; A28754.
DR PIR; B34493; B34493.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Cartilage; Signal.
FT SIGNAL 1 23
FT CHAIN 24 >416 COLLAGEN ALPHA 1(IX) CHAIN.
FT NON CONS 416 417 POTENTIAL.
FT CHAIN 24 503
FT DOMAIN 24 266 NONHELICAL REGION (N-TERMINAL) (NC4).
FT DISULFID 44 198 POTENTIAL.
FT DISULFID 242 252 POTENTIAL.
FT CONFLICT 25 25 Y -> I (IN REF. 2).
FT CONFLICT 51 51 E -> Q (IN REF. 2).
SQ SEQUENCE 503 AA: 51879 MW: DDC0808DC70A7C5 CRC64:

Query Match 27.5% Score 65.5; DB 1; Length 503;
Best Local Similarity 42.1%; Pred. No. 1.8;
Matches 16; Conservative 2; Mismatches 15; Indels 5; Gaps 1;

QY 2 PLGPKMPEYFGRGLASPGFGEYANDQERRMTLTAPG 39
| | | | | | | | | | | | | | | |
DB 298 PPGPGPDGAGKAGSGPLGEPGAD-----GLTGPDG 330

RESULT 19
TLD_DROME STANDARD; PRT; 1057 AA.
ID TLD_DROME
AC P25723; Q9VC46;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dorsal-ventral patterning tolloid protein precursor (BC 3.4.24.-).
GN TLD OR CG6688.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cancon-S;
RA Shinnell M.J., Ferguson E.L., Childs S.R., O'Connor M.B.;
RT "The Drosophila dorsal-ventral patterning gene tolloid is related to
human bone morphogenetic protein 1."
Cell 67:469-481(1991).
RN [2]
RP SEQUENCE FROM N.A.
MEDLINE=95324373; PubMed=7600963;
RA Finelli A.L., Bossis C.A., Xie T., Padgett R.W.;
RT "Mutational analysis of the Drosophila tolloid gene, a human BMP-1
homolog."
Development 120:861-870(1994).

```



DE	Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
GN	NRP2.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX	NCBI_Taxid=10116; [1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Sprague-Dawley;
RX	KJ062875.1; PubMed=9288754;
RA	Kojodkin A.L., Levensgood D.V., Rowe E.G., Tal Y.-T., Giger R.J.,
RA	Girty D.D.;
RT	"Neuropilin is a semaphorin II receptor."
RL	Cell 90:753-762(1997).
CC	-1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3E, VEGF-165 AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE CNS AND IN OTHER NONNEURAL TISSUES INCLUDING MESENCHYMAL TISSUE LINING IN THE RBFS.
CC	-1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> )
CC	-----
DR	EMBL, AF016297; AAC53338.1; -
DR	HSSP, P12259; 1CZT.
DR	InterPro: IPR000859; CUB_domain.
DR	InterPro: IPR000421; FA58_C.
DR	InterPro: IPR000998; MAM_domain.
DR	Pfam, PF00431; CUB; 2.
DR	Pfam, PF00629; MAM; 1.
DR	Pfam, PF00754; F5_F8_type_C; 2.
DR	SMART, SMO0042; CUB; 2.
DR	SMART, SMO0231; FA58C; 2.
DR	SMART, SMO0137; MAM; 1.
DR	PROSITE, PS01180; CUB; 2.
DR	PROSITE, PS01285; FA58C_1; 2.
DR	PROSITE, PS01286; FA58C_2; 2.
DR	PROSITE, PS50060; MAM_2; 1.
DR	Transmembrane; Glycoprotein;
FT	SIGNAL 1 22 Neurone; signal; Repeat; Receptor.
FT	CHAIN 1 22 POTENTIAL.
FT	CUB 1 22 NEUROPILIN-2.
FT	DOMAIN 23 925 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 859 883 POTENTIAL.
FT	DOMAIN 884 925 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 28 142 CUB 1.
FT	DOMAIN 149 267 CUB 2.
FT	DOMAIN 277 427 F5/F8 TYPE C 1.
FT	DOMAIN 434 592 F5/F8 TYPE C 2.
FT	DOMAIN 642 802 MAM.
FT	DISULFID 28 55 BY SIMILARITY.
FT	DISULFID 83 105 BY SIMILARITY.
FT	DISULFID 149 175 BY SIMILARITY.
FT	DISULFID 208 230 BY SIMILARITY.
FT	DISULFID 277 427 BY SIMILARITY.
FT	DISULFID 434 592 BY SIMILARITY.
FT	DISULFID 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 833 833 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
FO	SEQUENCE 925 AA; 103896 MW; 3BF62903FF644B51C CRC64;

Query Match	26.9%	Score 64	DB 1	Length 925
Best Local Similarity	33.3%	Pred. No. 5.1		
Matches	12	Conservative	6	Mismatches 18; Indels 0; Gaps 0;
QY	2	PLGPKMPEPVFGRLASPGGPGCIYANDQERRRLTLP	37	
DB	27	PCGGRLNSKDAGYITSPSPQDPYPSHQNCENWVYAP	62	

Search completed: January 11, 2003, 10:49:28  
Job time : 67 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 09:14:14 : Search time 117 Seconds  
(without alignments)  
46.695 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPLGPKWPEPVGRLASPGF.....GEVANDERRWTLTAPGVR 41

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database: A.Geneseq\_101002.\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	41	23	AAE14563 Human 20 kDa MASP-
2	238	100.0	671	23	AAE14565 Human mature MASP-
3	238	100.0	686	23	AAE14564 Human MASP-2 prote
4	238	100.0	686	23	AAE14568 Human MASP-2 prote
5	238	100.0	1034	22	ABG21134 Novel human diagno
6	93.5	39.3	174	22	AAU87257 Novel central nerv
7	93.5	39.3	174	22	ABU10255 Human CDNA SFQ ID
8	93.5	39.3	174	22	AAU19914 Novel human calciu
9	93	39.1	60	21	AAU01729 Human secreted pro
10	93	39.1	97	21	AAU00221 Human secreted pro

11	93	39.1	728	22	AAU47559	Protease PRS-1.
12	93	39.1	728	22	AAU85060	Human serine prote
13	91.5	38.4	525	22	AAU70539	Human PRO10 protei
14	91.5	38.4	525	22	AAU70541	Human PRO11 protei
15	91.5	38.4	525	22	AAU70542	Human PRO12 protei
16	91.5	38.4	525	22	AAU70543	Human SE26 mature
17	91.5	38.4	829	23	AAE15854	Human SE26 protein
18	91.5	38.4	853	23	AAE15853	Human SE26 protein
19	91.5	38.4	993	22	AAU70537	Human PRO7 protein
20	91.5	38.4	994	22	AAU70538	Human PRO7 protein
21	91.5	38.4	994	23	AAU81976	Human secreted pro
22	91	38.2	109	23	AAE15860	Human SE26 protein
23	84	35.3	3138	22	AAU61958	Drosophila melanog
24	83.5	35.1	1290	18	AAU07609	Rat von Ebner's gl
25	81	34.0	351	21	AAU58273	Lung cancer associ
26	81	34.0	449	23	AAU84299	Human endometrial
27	81	34.0	705	22	AAU50288	Human endometrial
28	81	34.0	705	22	AAU80757	Human C1r protein.
29	79	33.2	689	22	AAU00713	Human novel G-prot
30	79	33.2	690	22	AAU00208	Human novel G-prot
31	79	33.2	690	22	AAU00717	Human novel G-prot
32	79	33.2	717	22	AAU00196	Human novel G-prot
33	79	33.2	718	22	AAU00192	Human novel G-prot
34	79	33.2	718	22	AAU00200	Human novel G-prot
35	79	33.2	1083	22	AAU00714	Human novel G-prot
36	79	33.2	1084	22	AAU00209	Human novel G-prot
37	79	33.2	1084	22	AAU00718	Human novel G-prot
38	79	33.2	1111	22	AAU00197	Human novel G-prot
39	79	33.2	1112	22	AAU00193	Human novel G-prot
40	79	33.2	1112	22	AAU00201	Human novel G-prot
41	79	33.2	1192	22	AAU00712	Human novel G-prot
42	79	33.2	1192	22	AAU00719	Human novel G-prot
43	79	33.2	1193	22	AAU00207	Human novel G-prot
44	79	33.2	1193	22	AAU00716	Human novel G-prot
45	79	33.2	1220	22	AAU00195	Human novel G-prot
46	79	33.2	1221	22	AAU00191	Human novel G-prot
47	79	33.2	1221	22	AAU00199	Human novel G-prot
48	79	33.2	1221	22	AAU00210	Human novel G-prot
49	79	33.2	1222	22	AAU00205	Human novel G-prot
50	79	33.2	1222	22	AAU00715	Human novel G-prot
51	79	33.2	1249	22	AAU00194	Human novel G-prot
52	79	33.2	1250	22	AAU00196	Human novel G-prot
53	79	33.2	1250	22	AAU00198	Human novel G-prot
54	78	32.8	334	22	AAU75322	Gene 36 human secr
55	78	32.8	350	22	AAU93317	Human polypeptide,
56	78	32.8	449	18	AAU19720	Human smooth muscl
57	78	32.8	449	19	AAU56151	Procollagen C-pi
58	78	32.8	449	20	AAU24473	Human smooth muscl
59	78	32.8	458	21	AAU43837	Human cancer assoc
60	78	32.8	464	22	AAU93631	Human polypeptide,

## ALIGNMENTS

RESULT 1	AAE14563	standard: peptide; 41 AA.
ID	AAE14563	
AC	AAE14563	
XX		
XX		
DT	17-MAY-2002	(first entry)
XX		
XX		
DE	Human 20 kDa MASP-2 fragment.	
XX		
KW	Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;	
KW	complement fixation; infection; microbe; retrovirus; HIV; abortion;	
XX	human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.	
OS	Homo sapiens.	
XX		
XX		
PN	WO200206460-A2.	
XX		

PD	24-JAN-2002.
XX	
XX	13-JUL-2001.; 2001WO-DK00499.
PF	
PR	13-JUL-2000; 2000DK-0001089.
PR	01-JUN-2001; 2001DK-0000870.
XX	
PA	(JENS/) JENSENIUS J C.
PA	(THIE/) THIEL S.
PI	
PI	Jensenius JC, Thiel S;
XX	
DR	WPI; 2002-179791/23.
XX	
PT	Use of a polypeptide comprising amino acid sequence derived from
PT	mannan-binding lectin associated serine protease-2 (MASP-2) for
PT	producing pharmaceutical composition, to treat bacterial, fungal, viral
PT	infections
XX	
PT	Claim 39; Page 3; 76pp; English.
XX	
XX	The invention relates to use of a polypeptide derived from
XX	mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
XX	producing a pharmaceutical composition. MASP-2 is a complement-
XX	fixing enzyme and involved in lectin pathway of complement activation.
XX	The pharmaceutical composition comprising MASP-2 is useful for
XX	treating infectious caused by microbes such as fungus, yeast,
XX	retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
XX	bacteria which are resistant to at least one antibiotic medication or
XX	multiresistant. The polynucleotide encoding MASP-2 is useful for
XX	treating patients deficient in MASP-2. The invention also discloses
XX	MASP-2 assays which are useful for determination of MASP-2 activity or
XX	levels in patients suffering from e.g. infections, inflammatory disorders
XX	and spontaneous recurrent abortion. The pharmaceutical composition
XX	comprising MASP-2 inhibitor is useful for treating inflammatory
XX	disorders. The present sequence is human MASP-2 20 kda fragment.
XX	
XX	Sequence 41 AA:
XX	
XX	Query Match 100.0%; Score 238; DB 23; Length 41;
XX	Best Local Similarity 100.0%; Pred. No. 2.6e-23;
XX	Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	
QY	1 TPLGPKWPEPVVGRGLASPGFGEYVANDQERRWTLTAPPGYR 41
DB	1 TPLGPKWPEPVVGRGLASPGFGEYVANDQERRWTLTAPPGYR 41
XX	
XX	Sequence 865
XX	AAEI4565 standard; peptide; 671 AA.
XX	AAEI4565;
XX	
DT	17-MAY-2002 (first entry)
XX	
DE	Human mature MASP-2 protein.
XX	
KW	Human; mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KW	complement fixation; infection; microbe; retrovirus; HIV; abortion;
XX	human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
OS	
OS	Homo sapiens.
XX	
XX	Key Location/Qualifiers
XX	Region 15..671
XX	/note="This region is specifically claimed in
XX	claim 45"
XX	
XX	WO200206460-A2.
XX	
XX	24-JAN-2002.

[illegible]



FT Region 30..444  
 /note= "This region is specifically claimed in claim 41."  
 FT The region 30 to 296 is specifically claimed in claim 47."  
 FT 108..134  
 /note= "Fragment obtained by amino acid sequencing  
 of peptides"  
 FT Domain 135..146  
 /label= EGF-like\_domain  
 FT Region 138..296  
 /note= "This region is specifically claimed in claim 43"  
 FT Misc-difference 155  
 /note= "Encoded by CA"  
 FT Misc-difference 156  
 /note= "Encoded by C"  
 FT Domain 183..199  
 /label= C1r/C1s-like\_domain  
 FT Domain 293..307  
 /label= CCP-1\_domain  
 /note= "Complement control protein domain"  
 FT Domain 363..376  
 /label= CCP-2\_domain  
 FT Region 377..388  
 /note= "Fragment obtained by amino acid sequencing  
 of peptides"  
 FT Region 410..417  
 /note= "Fragment obtained by amino acid sequencing  
 of peptides"  
 FT Region 432..441  
 /note= "Linker"  
 FT Domain 445..468  
 /label= Serine\_protease\_domain  
 FT Active-site 483  
 FT Active-site 532  
 FT Active-site 633  
 PN WO200206460-A2.  
 PD 24-JAN-2002.  
 PF 13-JUL-2001; 2001WO-DK00499.  
 PR 13-JUL-2000; 2000DK-0001089.  
 PR 01-JUN-2001; 2001DK-0000870.  
 PA (JENSEN/) JENSENIUS J C.  
 PA (THIE/) THIEL S.  
 PA Jensenius JC, Thiel S;  
 WIPI: 2002-179791/23.  
 DR N-PSDB: AAD24224.  
 XX Use of a polypeptide comprising amino acid sequence derived from  
 PT mannan-binding lectin associated serine protease-2 (MASP-2) for  
 PT producing pharmaceutical composition, to treat bacterial, fungal, viral  
 PT infections -  
 XX Claim 41: Fig 6; 76pp: English.  
 PS The invention relates to use of a polypeptide derived from  
 XX mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for  
 CC producing a pharmaceutical composition. MASP-2 is a complement-  
 CC fixing enzyme and involved in lectin pathway of complement activation.  
 CC The pharmaceutical composition comprising MASP-2 is useful for  
 CC treating infections caused by microbes such as fungus, yeast,  
 CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic  
 CC bacteria which are resistant to at least one antibiotic medicament or  
 CC multiresistant. The polynucleotide encoding MASP-2 is useful for  
 CC treating patients deficient in MASP-2. The invention also discloses  
 CC MASP-2 assays which are useful for determination of MASP-2 activity or  
 CC levels in patients suffering from e.g. infections, inflammatory disorders  
 CC and spontaneous recurrent abortion. The pharmaceutical composition  
 CC comprising MASP-2 inhibitor is useful for treating inflammatory

CC disorders. The present sequence is human MASP-2 protein.  
 CC Note: The present sequence is stated as being the same as  
 CC SEQ ID NO:2 shown in sequence listing of the specification (AAE14568).  
 CC However the sequences differ at various locations.  
 XX  
 SQ Sequence 686 AA:  
 Query Match 100.0%; Score 238; DB 23;  
 Best Local Similarity 100.0%; Pred. No.5.5e-22;  
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TPLGPKWPEPVFGRLASPGFGEXANDQERRMTLTAPGGR 41  
 Db 16 TPLGPKWPEPVFGRLASPGFGEXANDQERRMTLTAPGGR 56  
 RESULT 4  
 AAE14568  
 ID AAE14568 standard; Protein; 686 AA.  
 AC AAE14568;  
 DT 17-MAY-2002 (first entry)  
 DE Human MASP-2 protein, alternative version.  
 XX  
 KW Human; mannan-binding lectin associated serine protease-2; MASP-2; MBL;  
 KW complement fixation; infection; microbe; retrovirus; HIV; abortion;  
 KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..15  
 FT Protein /label= Signal\_peptide  
 FT 16..686  
 FT /label= Mature\_MASP-2\_protein  
 FT Misc-difference 155  
 /note= "Encoded by CA"  
 FT Misc-difference 156  
 /note= "Encoded by C"  
 FT Misc-difference 298  
 /note= "Encoded by CAG"  
 FT Misc-difference 299  
 /note= "Encoded by CCT"  
 XX  
 PN WO200206460-A2.  
 PD 24-JAN-2002.  
 PF 13-JUL-2001; 2001WO-DK00499.  
 PR 13-JUL-2000; 2000DK-0001089.  
 PR 01-JUN-2001; 2001DK-0000870.  
 PA (JENSEN/) JENSENIUS J C.  
 PA (THIE/) THIEL S.  
 PA Jensenius JC, Thiel S;  
 WIPI: 2002-179791/23.  
 DR N-PSDB: AAD24224.  
 XX Use of a polypeptide comprising amino acid sequence derived from  
 PT mannan-binding lectin associated serine protease-2 (MASP-2) for  
 PT producing pharmaceutical composition, to treat bacterial, fungal, viral  
 PT infections -  
 XX Claim 41: Page 71-73; 76pp: English.  
 PS The invention relates to use of a polypeptide derived from  
 XX mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for  
 CC producing a pharmaceutical composition. MASP-2 is a complement-

CC fixing enzyme and involved in lectin pathway of complement activation.  
CC The pharmaceutical composition comprising MASP-2 is useful for  
CC treating infections caused by microbes such as fungus, yeast,  
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic  
CC bacteria which are resistant to at least one antibiotic medication or  
CC multiresistant. The polynucleotide encoding MASP-2 is useful for  
CC treating patients deficient in MASP-2. The invention also discloses  
CC MASP-2 assays which are useful for determination of MASP-2 activity or  
CC levels in patients suffering from e.g. infections, inflammatory disorders  
CC and spontaneous recurrent abortion. The pharmaceutical composition  
CC comprising MASP-2 inhibitor is useful for treating inflammatory  
CC disorders. The present sequence is human MASP-2 protein.  
CC Note: The present sequence is stated as being the same as  
CC SEQ ID NO:2 shown in figure 6 of the specification (AAM14564).  
CC However the sequences differ at various locations.  
XX  
SQ Sequence 686 AA:  
Query Match 100.0%; Score 238; DB 23; Length 686;  
Best Local Similarity 100.0%; Pred. No. 5.5e-22;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db. 1 TPLGPKWPEPVFGRLASPGPGGYANDQERRWTLTAPGVR 41  
16 TPLGPKWPEPVFGRLASPGPGGYANDQERRWTLTAPGVR 56  
RESULT 5  
ABG21134  
ID ABG21134 standard; Protein: 1034 AA.  
XX  
AC ABG21134:  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #21125.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO2001/5067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PE 31-MAR-2000; 2000US-0540217. } *the lute*  
33-AUG-2000; 2000US-0649167.  
XX  
(HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS85321.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20; SEQ ID NO 51493; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1034 AA:  
Query Match 100.0%; Score 238; DB 22; Length 1034;  
Best Local Similarity 100.0%; Pred. No. 8.7e-22;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db. 1 TPLGPKWPEPVFGRLASPGPGGYANDQERRWTLTAPGVR 41  
752 TPLGPKWPEPVFGRLASPGPGGYANDQERRWTLTAPGVR 792  
RESULT 6  
AAU87257  
ID AAU87257 standard; Protein: 174 AA.  
XX  
AC AAU87257:  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Novel central nervous system protein #167.  
XX  
XX  
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
KW myocardial infarction; wound healing; cell proliferation; skin aging;  
KW food additive; food preservative; gene therapy.  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200155318-A2.  
XX  
PD 02-AUG-2001.  
XX  
XX  
PE 17-JAN-2001; 2001WO-US01332.  
XX  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226779.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246538.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0250393.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251857.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI: 2001-581633/65.  
N-PSDB; ABK43587.  
  
New isolated nucleic acid encoding a protein for diagnosing,  
preventing, treating or ameliorating medical conditions and used as  
food additives or preservatives -  
  
Claim 9; SEQ ID No 775; 837pp; English.  
  
The invention describes an isolated nucleic acid molecule (I) encoding a  
novel central nervous system protein. (I) and polypeptides (II) encoded  
by (I), are used to treat a medical conditions and in diagnosis of a  
pathological condition. Disorders which are diagnosed or treated include  
autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,  
CC angio genesis, nervous system disorders e.g. Alzheimer's disease and  
CC amyloidotic lateral sclerosis, infections caused by bacteria, viruses  
e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
CC adenocarcinomas and irritable bowel syndrome, reproductive system  
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
leukemia, disorders involving neovascularisation e.g. malignancies.

PR	22-AUG-2000	2000US-226866P
PR	22-AUG-2000	2000US-227188P
PR	23-AUG-2000	2000US-227109P
PR	30-AUG-2000	2000US-228924P
PR	01-SEP-2000	2000US-229287P
PR	01-SEP-2000	2000US-229343P
PR	01-SEP-2000	2000US-229344P
PR	01-SEP-2000	2000US-229345P
PR	05-SEP-2000	2000US-229509P
PR	05-SEP-2000	2000US-229513P
PR	06-SEP-2000	2000US-230437P
PR	06-SEP-2000	2000US-230438P
PR	08-SEP-2000	2000US-231242P
PR	08-SEP-2000	2000US-231244P
PR	08-SEP-2000	2000US-231245P
PR	08-SEP-2000	2000US-231413P
PR	08-SEP-2000	2000US-231414P
PR	08-SEP-2000	2000US-232081P
PR	12-SEP-2000	2000US-233018P
PR	12-SEP-2000	2000US-233196P
PR	14-SEP-2000	2000US-233397P
PR	14-SEP-2000	2000US-233398P
PR	14-SEP-2000	2000US-233399P
PR	14-SEP-2000	2000US-233400P
PR	14-SEP-2000	2000US-233401P
PR	14-SEP-2000	2000US-233063P
PR	14-SEP-2000	2000US-233064P
PR	14-SEP-2000	2000US-233065P
PR	21-SEP-2000	2000US-233422P
PR	21-SEP-2000	2000US-234272P
PR	25-SEP-2000	2000US-234397P
PR	25-SEP-2000	2000US-234398P
PR	26-SEP-2000	2000US-235348P
PR	27-SEP-2000	2000US-235344P
PR	27-SEP-2000	2000US-235836P
PR	29-SEP-2000	2000US-236367P
PR	29-SEP-2000	2000US-236368P
PR	29-SEP-2000	2000US-236369P
PR	29-SEP-2000	2000US-236370P
PR	02-OCT-2000	2000US-236802P
PR	02-OCT-2000	2000US-237037P
PR	02-OCT-2000	2000US-237038P
PR	02-OCT-2000	2000US-237039P
PR	02-OCT-2000	2000US-237040P
PR	13-OCT-2000	2000US-239357P
PR	13-OCT-2000	2000US-239357P
PR	20-OCT-2000	2000US-240960P
PR	20-OCT-2000	2000US-241122P
PR	20-OCT-2000	2000US-241175P
PR	20-OCT-2000	2000US-241176P
PR	20-OCT-2000	2000US-241787P
PR	20-OCT-2000	2000US-241809P
PR	20-OCT-2000	2000US-241816P
PR	20-OCT-2000	2000US-242452P
PR	01-NOV-2000	2000US-244617P
PR	01-NOV-2000	2000US-244618P
PR	08-NOV-2000	2000US-246475P
PR	08-NOV-2000	2000US-246476P
PR	08-NOV-2000	2000US-246477P
PR	08-NOV-2000	2000US-246523P
PR	08-NOV-2000	2000US-246524P
PR	08-NOV-2000	2000US-246525P
PR	08-NOV-2000	2000US-246526P
PR	08-NOV-2000	2000US-246527P
PR	08-NOV-2000	2000US-246552P
PR	08-NOV-2000	2000US-246553P
PR	08-NOV-2000	2000US-246609P
PR	08-NOV-2000	2000US-246610P
PR	08-NOV-2000	2000US-246611P
PR	08-NOV-2000	2000US-246613P
PR	17-NOV-2000	2000US-249207P





DR WPI: 2000-500381/45.  
DR N-PSDB: AAC01735.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PR diagnostic, forensic, gene therapy and chromosome mapping procedures -  
PS  
XX

Claim 13: SEQ ID 5810; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number  
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
CC different tissues. EST sequences usually correspond mainly to the 3'  
CC untranslated region (UTR) of the mRNA because they are often obtained  
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
CC those cases where longer cDNA sequences have been obtained, the full 5'  
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
CC 3' ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream  
CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 60 AA:  
SQ

Query Match 39.1%; Score 93; DB 21; Length 60;  
Best Local Similarity 45.2%; Pred. NO. 0.00014;  
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

OY 11 VEGRLASPGFPEGVANDGERRWLTAPPGCYR 41  
          ::: |||: | : | : | : | : | :  
.DB 1 MFQGIOSPGYPDSYSDSEVTWNTVDPGFR 31

RESULT 10  
AAAG00221  
ID AAG00221 standard; Protein: 97 AA.

XX AAG00221;  
XX  
XX AAG00221;  
DT 06-OCT-2000 (first entry)  
DE  
DE Human secreted protein, SEQ ID NO: 4302.  
XX  
XX - Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping.  
XX Homo sapiens.  
● EPI033401-A2.

XX PD 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
XX 26-FEB-1999; 99US-0122487.  
XX  
XX (GEST ) GENSET.  
PA  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
PI  
XX WPI: 2000-500381/45.  
XX  
DR N-PSDB: AAC00227.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PR diagnostic, forensic, gene therapy and chromosome mapping procedures -  
PS  
XX Claim 13: SEQ ID 4302; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number  
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
CC different tissues. EST sequences usually correspond mainly to the 3'  
CC untranslated region (UTR) of the mRNA because they are often obtained  
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
CC those cases where longer cDNA sequences have been obtained, the full 5'  
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
CC 3' ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream  
CC regulatory sequences and to design expression and secretion vectors.

[illegible]

FT		Modified-site	/label= Potential phosphorylation site
FT		361	
FT		/label= Potential phosphorylation site	
FT		381	
FT		/label= Potential phosphorylation site	
FT		385	
FT		/label= Potential glycosylation site	
FT		388	
FT		/label= Potential phosphorylation site	
FT		396	
FT		/label= Potential phosphorylation site	
FT		407	
FT		/label= Potential glycosylation site	
FT		472	
FT		/label= Potential phosphorylation site	
FT		476..531	
FT		/label= Serine protease active site	
FT		501	
FT		/label= Potential phosphorylation site	
FT		506	
FT		/label= Potential phosphorylation site	
FT		533	
FT		/label= Potential glycosylation site	
FT		599	
FT		/label= Potential glycosylation site	
FT		601	
FT		/label= Potential phosphorylation site	
FT		612	
FT		/label= Potential phosphorylation site	
FT		619	
FT		/label= Potential phosphorylation site	
FT		632	
FT		/label= Potential phosphorylation site	
FT		640	
FT		/label= Potential glycosylation site	
FT		642	
FT		/label= Potential phosphorylation site	
FT		643..692	
FT		/label= Serine protease active site	
FT		675	
FT		/label= Potential phosphorylation site	
FT		720	
FT		/label= Potential phosphorylation site	
PN	WO200171004-A2.		
XX			
XX			
PD	27-SEP-2001.		
XX			
XX	16-MAR-2001; 2001WO-US08441.		
PR	27-MAR-2000; 2000US-190708P.		
PR	30-MAR-2000; 2000US-193182P.		
PR	14-APR-2000; 2000US-197086P.		
PR	20-APR-2000; 2000US-199022P.		
PR	28-APR-2000; 2000US-200227P.		
XX			
PA	(INCYTE GENOMICS INC.		
PI	Yue H., Lu DAM, Pollicy JL, Delegeane AM, Tribouley CM, Khan FA;		
PI	Au-Yang J, Bandman O, Lal P, Borowsky ML, Gandhi AR, Hillman JL;		
PI	Tang YT, Burford N, Baugh MR, Nguyen DB, Yao MG, Walia NK, He A;		
PI	Hatalia A, Lu Y, Paterson C;		
XX			
DR	WPI: 2001-611509/70.		
DR	N-PSDB; AAH43512.		
XX			
PT	New polypeptide for treating gastrointestinal, cardiovascular and		
PT	autoimmune disorders, comprises novel human proteases (PRTS) and		
PT	polynucleotides -		
XX			
PS	Claim 1, Page 109-11; 12pp; English.		
CC	The sequences given in AAB47559-69 show novel human proteases PRTS-1		





DB 403 PFWDSEPCVIAACGGVIRNGTGRIVSPGPGNYSNLTCHMLLEAPESQR 454

# RESULT 15

ID AAB70541 standard; Protein; 525 AA.

AC AAB70541;

DT 09-MAY-2001 (first entry)

DE Human PRO11 protein sequence SEQ ID NO:22.

XX Human; PRO; cytosolic; immunomodulatory; reproduction;

KW gene therapy; cell proliferation; differentiation disorder; cancer;

XX immune associated disorder; gestational disease; pre-clampsia.

OS Homo sapiens.

XX WO200110902-A2.

PF 11-AUG-2000; 2000WO-US21857.

XX 11-AUG-1999; 99US-0148433.

PR 10-AUG-2000; 2000US-0148433.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Fernandes E;

DR WPI: 2001-147509/15.

XX N-PSDB; AAF74442.

PT Nucleic acids encoding secreted polypeptides, designated PROX

XX polypeptides, useful for treating a syndrome associated with a

PS PROX-associated disorder, e.g. cancer -

XX Claim 1; Page 38-40; 166pp; English.

CC The present invention describes isolated nucleic acids encoding secreted

CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where

CC X is an integer from 1 to 17). PROX polypeptides have cytosolic,

CC immunomodulatory and reproduction activities, and can be used in gene

CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,

CC nucleic acids and antibodies are useful in the manufacture of a

CC medicament for treating a syndrome associated with a PROX-associated

CC disorder, e.g. a cell proliferation and/or differentiation disorder

CC (e.g. cancer or immune associated disorders) and a gestational disease

CC activity or of latency or predisposition to a PROX-associated disorder.

CC AAF74432 to AAF74448 encode the specifically claimed human PROX

CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.

XX Sequence 525 AA;

SO Query Match 38.4%; Score 91.5; DB 22; Length 525;

Best Local Similarity 40.4%; Pred. No. 0.0023;

Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

QY 5 PKW--PEPV-----FGRLASGPFGEYANDERRRWTLPAPGYR 41

DB 403 PFWDSEPCVIAACGGVIRNGTGRIVSPGPGNYSNLTCHMLLEAPESQR 454

RESULT 16

AAB70542

ID AAB70542 standard; Protein; 526 AA.

AC AAB70542;

XX 09-MAY-2001 (first entry)

DE Human PRO12 protein sequence SEQ ID NO:24.

XX Human; PRO; cytosolic; immunomodulatory; reproduction;

KW gene therapy; cell proliferation; differentiation disorder; cancer;

XX immune associated disorder; gestational disease; pre-clampsia.

OS Homo sapiens.

XX WO200110902-A2.

PN 15-FEB-2001.

PD 11-AUG-2000; 2000WO-US21857.

PF 11-AUG-1999; 99US-0148433.

PR 10-AUG-2000; 2000US-0148433.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Fernandes E;

DR WPI: 2001-147509/15.

XX N-PSDB; AAF74443.

PT Nucleic acids encoding secreted polypeptides, designated PROX

XX polypeptides, useful for treating a syndrome associated with a

PS PROX-associated disorder, e.g. cancer -

XX Claim 1; Page 41-43; 166pp; English.

CC The present invention describes isolated nucleic acids encoding secreted

CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where

CC X is an integer from 1 to 17). PROX polypeptides have cytosolic,

CC immunomodulatory and reproduction activities, and can be used in gene

CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,

CC nucleic acids and antibodies are useful in the manufacture of a

CC medicament for treating a syndrome associated with a PROX-associated

CC disorder, e.g. a cell proliferation and/or differentiation disorder

CC (e.g. cancer or immune associated disorders) and a gestational disease

CC activity or of latency or predisposition to a PROX-associated disorder.

CC AAF74432 to AAF74448 encode the specifically claimed human PROX

CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.

XX Sequence 526 AA;

SO Query Match 38.4%; Score 91.5; DB 22; Length 526;

Best Local Similarity 40.4%; Pred. No. 0.0023;

Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

QY 5 PKW--PEPV-----FGRLASGPFGEYANDERRRWTLPAPGYR 41

DB 403 PFWDSEPCVIAACGGVIRNGTGRIVSPGPGNYSNLTCHMLLEAPESQR 454

RESULT 17

AAE15854

ID AAE15854 standard; Protein; 829 AA.

AC AAE15854;

XX 26-MAR-2002 (first entry)

DE Human SEZ6 mature protein.

XX Human; SEZ6; neural regeneration; seizure; infertility; gene therapy;

KW stroke; Alzheimer's disease; Huntington's disease; myasthenia gravis;

KW neurotrophic; neuroprotective; anticonvulsant; cerebroprotective; stroke;

KW vasotrophic; neuronal growth; growth factor-mediated chemotaxis; trauma;

KW neurological disorder; muscular dystrophy; muscle injury; vulvectomy;

KW amyotrophic lateral sclerosis; multiple sclerosis; ischemia; diabetes;

KW epilepsy; Parkinson's disease; sexual development; gene mapping;

KW	impotence; libido.	
XX		
OS	Homo sapiens.	
PN	WO200183552-A2.	
PD	08-NOV-2001.	
PF	17-APR-2001; 2001WO-US10809.	
PR	28-APR-2000; 2000US-200200P.	
PA	(ELIL ) LILLY & CO ELI.	
PI	Su EW;	
DR	WPI: 2002-082840/11.	
DR	N-PSDB: AAD25344.	
XX		
XX		
XX		
PS	Claim 9; Page 122-125; 127pp; English.	
XX		
CC	The invention relates to an isolated human SEZ6 (hSEZ6) polypeptide and	
CC	its corresponding DNA molecule. SEZ6 is useful for enhancing neuronal	
CC	growth, neurite outgrowth, neuronal regeneration, neuronal survival,	
CC	growth factor-mediated chemotaxis, altered behavioural patterns, e.g.,	
CC	sleep or eating disorders and for treating neurological disorders such	
CC	as trigeminal neuralgia, Bell's palsy, myasthenia gravis, muscular	
CC	dystrophy, muscle injury, intervertebral disk syndrome, thoracic outlet	
CC	destruction syndrome, amyotrophic lateral sclerosis, multiple sclerosis,	
CC	ischaemia associated with stroke, neuropathy associated with diabetes,	
CC	spinal cord trauma, facial nerve crush and other trauma, Huntington's	
CC	disease. SEZ6 is useful for treating a patient suffering from a	
CC	neurological disorder such as epilepsy, Alzheimer's disease, Parkinson's	
CC	disease, seizure related disorder or a disorder associated with stroke.	
CC	SEZ6 DNA is useful as probes for gene mapping and for detecting	
CC	transcription, translation and/or expression of hse26 polypeptide in	
CC	human tissue. A transgenic animal is useful as an animal model in	
CC	research and drug development procedures, and for testing compounds or	
CC	other treatment modalities which may prevent, suppress or cure a	
CC	pathology or disease associated with hse26 activities. SEZ6 is also	
CC	useful for treating abnormal primary or secondary sexual development,	
CC	e.g., impotence, infertility or reduced libido. The hse26 polynucleotide	
CC	is useful for treating the above mentioned disorders by gene therapy	
CC	techniques. The present sequence is human SEZ6 mature protein.	
SO	Sequence 829 AA;	
Query Match	38.4%; Score 91.5; DB 23; Length 829;	
Best Local Similarity	40.4%; Pred. No. 0.0038;	
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2		
OY	5 PKW--PEPV-----FGRLASPGFGEYANDQERWTTAPGGR 41	
DB	379 PFWDSKEPVCIACGVTIRNATTGRIVSGFGPNVSNNTTCMHLLEAPGGR 430	
RESULT 18		
AEI15853		
ID	AEI15853 standard; Protein: 853 AA.	
AC	AEI15853;	
DT	26-MAR-2002 (first entry)	
XX		
DE	Human SEZ6 protein.	
KW	Human; SEZ6: neural regeneration; seizure; infertility; gene therapy;	
KW	stroke; Alzheimer's disease; Huntington's disease; myasthenia gravis;	

KW	noctropic; neuroprotective; anticonvulsant; cerebroprotective; stroke;
KW	vasotropic; neuronal growth; growth factor-mediated chemotaxis; trauma;
KW	neurological disorder; muscular dystrophy; muscle injury; vulnerability;
KW	amyotrophic lateral sclerosis; multiple sclerosis; ischaemia; diabetes;
KW	epilepsy; Parkinson's disease; sexual development; gene mapping;
KW	impotence; libido.
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..24
FT	/label= Signal_peptide
FT	Protein 25..853
FT	/label= Mature_SEZ6_protein
XX	
PN	WO200183552-A2.
XP	
PD	08-NOV-2001.
XX	
PF	17-APR-2001; 2001WO-US10809.
XX	
PR	28-APR-2000; 2000US-200200P.
XX	
PA	(ELIL ) LILLY & CO ELI.
PX	
P1	Su EW;
XX	
DR	WPI: 2002-082840/11.
DR	N-PSDB: AAD25344.
XX	
PT	Novel human SEZ6 polypeptide useful for inducing neural regeneration,
PT	inhibiting neural degeneration, preventing seizures, and for treating
PT	infertility, Alzheimer's disease, stroke, seizures, Huntington's
PT	disease
XX	
PS	Claim 9; Page 119-122; 127pp; English.
XX	
CC	The invention relates to an isolated human SEZ6 (hSEZ6) polypeptide and
CC	its corresponding DNA molecule. SEZ6 is useful for enhancing neuronal
CC	growth, neurite outgrowth, neuronal regeneration, neuronal survival,
CC	growth factor-mediated chemotaxis, altered behavioural patterns, e.g.,
CC	sleep or eating disorders and for treating neurological disorders such
CC	as trigeminal neuralgia, Bell's palsy, myasthenia gravis, muscular
CC	dystrophy, muscle injury, intervertebral disk syndrome, thoracic outlet
CC	destruction syndrome, amyotrophic lateral sclerosis, multiple sclerosis,
CC	ischaemia associated with stroke, neuropathy associated with diabetes,
CC	spinal cord trauma, facial nerve crush and other trauma, Huntington's
CC	disease. SEZ6 is useful for treating a patient suffering from a
CC	neurological disorder such as epilepsy, Alzheimer's disease, Parkinson's
CC	disease, seizure related disorder or a disorder associated with stroke.
CC	SEZ6 DNA is useful as probes for gene mapping and for detecting
CC	transcription, translation and/or expression of hSEZ6 polypeptide in
CC	human tissue. A transgenic animal is useful as an animal model in
CC	research and drug development procedures, and for testing compounds or
CC	other treatment modalities which may prevent, suppress or cure a
CC	pathology or disease associated with hSEZ6 activities. SEZ6 is also
CC	useful for treating abnormal primary or secondary sexual development,
CC	e.g., impotence, infertility or reduced libido. The hSEZ6 polynucleotide
CC	is useful for treating the above mentioned disorders by gene therapy
CC	techniques. The present sequence is human SEZ6 protein.
XX	
SO	Sequence 853 AA:
Query Match	38.4%; Score 91.5; DB 23; Length 853;
Best Local Similarity	40.4%; Pred. NO. 0.0039;
Matches 21; Conservative	3; Mismatches 13; Indels 15; Gaps 2;
OY	5 PKW--PEPV-----FGLASPGFGGEYANDQERMTLTAPGGR 41   :       : :
Db	403 PFMDSKEPYCIAAGGVIRNATTGTIVSGFCGNYSNNLTGMLELAPGOR 454



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: January 11, 2003, 11:06:49 : Search time 3185 Seconds

(without alignments)  
374.636 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPLGKMPDPVFGRLASPGF.....GEYANDQERRMTLTAPPCVR 41

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched:

2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Command line parameters:

-MODE=frame\_plus.model -DEV=rlp  
-Q/cgn2\_l/uspto.spool/us09874198/runat\_10012003\_092043\_2562/app-query.fasta.1.199  
-DB=GenEmbl -QMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=60  
-DOCCALIGN=200 -THR.SCORE=dot -THR.MAX=100 -THR.MIN=0 -ALIGN=20 -MODE=LOCAL  
-OUTFMT=pcr -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=us09874198.accn.1.1.3637 -runat\_10012003\_092043\_2562 -NCRP=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEBUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGABEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_ntg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_of:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pin:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	238	100.0	683	9 HOSAI8282	Y18282 Homo sapien
2	238	100.0	725	9 AB008047	AB008047 Homo sapi
3	238	100.0	729	9 HOSAI8283	Y18283 Homo sapien
4	238	100.0	736	9 HOSAI8281	Y18281 Homo sapien
5	238	100.0	1184	9 HOSAI8284	Y18284 Homo sapien
6	238	100.0	2450	9 HSMAS2	X98400 H.sapiens m
7	238	100.0	2455	9 HSMAS2PR	Y09926 Homo sapien
8	238	100.0	2819	9 AB033742	AB033742 Homo sapi
9	238	100.0	4464	9 HOSAI8286	Y18286 Homo sapien
10	238	100.0	4465	9 HOSAI8287	Y18287 Homo sapien
11	238	100.0	4900	9 AF321558S1	AF321558 Homo sapl
12	238	100.0	112769	9 HSJ635E18	AL109811 Human DNA
13	238	100.0	733	9 MMU250369	AJ7250369 Mus muscu
14	183	76.9	817	10 MMU19160	Y19160 Mus musculu
15	183	76.9	1145	10 MMU19163	Y19163 Mus musculu
16	183	76.9	2050	10 AB009459	AB009459 Mus muscu
17	183	76.9	3070	10 BC013893	BC013893 Mus muscu
18	183	76.9	205573	10 AL591032	AL591032 Mouse DNA
19	183	76.9	224312	2 AL606969	AL606969 Mus muscu
20	180	75.6	2037	10 RNO277747	AJ277747 Rattus no
21	180	75.6	172473	2 AC115553	AC115553 Rattus no
22	180	75.6	186184	2 AC094176	AC094176 Rattus no
23	170	71.4	1800	10 MMU19161	Y19161 Rattus norv
24	164	68.9	658	10 RNO18570	Y18570 Rattus Norv
25	164	68.9	680	10 RNO18570	Y18570 Rattus Norv
26	164	68.9	715	10 RNO18568	Y18568 Rattus Norv
27	100	42.0	2282	5 AB009072	AB009072 Xenopus l
28	93	39.1	449	9 AB01082502	AB010813 Homo sapl
29	93	39.1	636	10 AF004661	AF004661 Rattus no
30	93	39.1	1013	9 AB00760502	AB007603 Homo sapl
31	93	39.1	2108	10 RNO277423	AJ277423 Rattus no
32	93	39.1	2787	9 HUMMASP	D28593 Human mRNA
33	93	39.1	3736	10 AB049755	AB049755 Mus muscu
34	93	39.1	3863	6 AX254423	AX254423 Sequence
35	93	39.1	3895	6 AX156466	AX156466 Sequence
36	93	39.1	3895	9 AF284421	AF284421 Homo sapl
37	93	39.1	4489	9 D17525	D17525 Human mRNA
38	93	39.1	5135	10 MUSCRARF	D16492 Mouse mRNA
39	93	39.1	62047	2 AC117672	AC117672 Mus muscu
40	93	39.1	154160	2 AC046154	AC046154 Homo sapl
41	93	39.1	173534	9 AC007920	AC007920 Homo sapl
42	93	39.1	207841	2 AC072019	AC072019 Homo sapl
43	91.5	38.4	1988	6 AX084227	AX084227 Sequence
44	91.5	38.4	2095	9 AF502130	AF502130 Homo sapl
45	91.5	38.4	2127	6 AX084223	AX084223 Sequence
46	91.5	38.4	2127	6 AX084225	AX084225 Sequence
47	91.5	38.4	2143	6 AX084229	AX084229 Sequence
48	91.5	38.4	2559	6 AX327363	AX327363 Sequence
49	91.5	38.4	3470	9 AK096473	AK096473 Homo sapl
50	91.5	38.4	3828	9 AK091522	AK091522 Homo sapl

51 91.5 38.4 3863 6 AX084219 Sequence  
 52 91.5 38.4 3879 6 AX084221 Sequence  
 53 91.5 38.4 4034 6 AX402474 Sequence  
 54 91.5 38.4 4047 9 HSM805489 AL834405 Homo sapi  
 55 91.5 38.4 4198 6 AX327362 Sequence  
 56 91.5 38.4 4227 9 AY038048 Homo sapi  
 57 91.5 38.4 4306 9 AF502129 Homo sapi  
 58 91 38.2 174701 2 AC024267 Homo sapi  
 59 36.6 220638 10 AL645723 Mouse DNA  
 60 85 35.7 33556 3 U00036 Caenorhabdit

## ALIGNMENTS

RESULT 1  
 HOSA18282 683 bp mRNA linear PRI 12-JUL-1999  
 LOCUS Homo sapiens mRNA for mannose binding lectin-associated serine  
 DEFINITION protease-2, alternatively spliced transcript (clone ph1-7).

Y18282  
 Y18282.1 GI:5459316  
 lectin-associated protein; mannose binding lectin-associated  
 protein; mannose binding protein; serine protease.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Storer,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorrup-Jensen,T.,  
 Jønsenius,O.C. and Schwaebie,W.J.  
 TITLE Two constituents of the initiation complex of the mannan-binding  
 JOURNAL lectin activation pathway of complement are encoded by a single  
 structural gene  
 MEDLINE J. Immunol. 162 (6), 3481-3490 (1999)  
 PUBMED 99192764

REFERENCE  
 AUTHORS 2 (bases 1 to 683)  
 TITLE Schwaebie,W.J.  
 JOURNAL Direct Submission  
 Submitted (19-OCT-1998) W.J. Schwaebie, Department of Microbiology  
 and Immunology, University of Leicester, University Road, PO Box  
 138, Leicester LE1 9HN, UK

COMMENT  
 FEATURES  
 source Location/Qualifiers  
 1..683  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1p36.2-3"  
 /clone="ph1-7"  
 /tissue\_type="liver"  
 1..683  
 /gene="MASP-2"  
 <1..528  
 /gene="MASP-2"  
 /note="alternative"  
 /codon\_start=1  
 /product="mannose binding lectin-associated serine  
 protease-2 related protein, MAP19 (19kDa)"  
 /protein\_id="CAB50729.1"  
 /db\_xref="GI:5459317"  
 /translation="CGSVATPLGPKWPEVFGRLASPGFGEVANDQERRWTLTAAPG  
 RRLVYFTHDELSHLCYDVKLSSGAKVLATCGDSTDERAPGKDFYSLGSS  
 LITRRSDYSNKKPFTGFAFYAAEDIDECQVAPGFAFPCDHCHNHLCGFCSCNAG  
 VYLHRRKTCSEQSL"  
 <1..15  
 /gene="MASP-2"  
 16..525  
 /gene="MASP-2"  
 /product="mannose binding lectin-associated serine  
 protease-2 related protein, MAP19 (19kDa)"  
 /note="alternative"  
 529..683

sig\_peptide  
 mat\_peptide  
 3'UTR

polyA\_signal  
 BASE COUNT 136 a 246 c 187 g 114 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6,45e-16 Length: 683  
 Score: 238.00 Matches: 41  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x HOSA18282 (1-683)

QY 1 ThProLeuGlyProIySTrProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
 |||  
 Db 16 ACCCCCTTAGCCCCGAGAGGCGCTGAACCTGTGTGGGGCGCTGCATCCCCGGCTTT 75

QY 21 ProGlyGluTrAlaAsnAspGlnGluArgTrPheLeuThrAlaProGlyTyr 40  
 |||  
 Db 76 CCAAGGAGATGATCCATGACCAAGAGCGGCGCTGACCTGACTACACCCCGGCTTAC 135

QY 41 Arg 41  
 |||  
 Db 136 CGC 138

RESULT 2  
 AB008047 725 bp mRNA linear PRI 08-JUN-1999  
 LOCUS Homo sapiens sMAP mRNA for small MBL-associated protein, complete  
 DEFINITION cds.

ACCESSION AB008047  
 VERSION AB008047.1 GI:5002493  
 KEYWORDS Small MBL-associated protein; sMAP.  
 SOURCE Homo sapiens cDNA to mRNA.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 JOURNAL 1 (sites)

Takahashi,M., Endo,Y., Fujita,T. and Matsushita,M.  
 A truncated form of mannose-binding lectin-associated serine  
 protease (MASP)-2 expressed by alternative polyadenylation is a  
 component of the lectin complement pathway  
 Int. Immunol. 11 (5), 859-863 (1999)

JOURNAL  
 MEDLINE 99262288  
 REFERENCE 2 (bases 1 to 725)  
 AUTHORS Takahashi,M., Matsushita,M. and Fujita,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-OCT-1997) Minoru Takahashi, Fukushima Medical  
 College, Department of Biochemistry, 1 Hikarigaoka, Fukushima,  
 Fukushima 960-12, Japan (E-mail:munolta@cc.fmu.ac.jp,  
 Tel:81-245-48-2111, Fax:81-245-48-2111)

COMMENT  
 FEATURES  
 source Location/Qualifiers  
 1..725  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..725  
 /gene="sMAP"  
 27..584  
 /gene="sMAP"  
 /product="small MBL-associated protein"  
 /protein\_id="BAA78616.1"  
 /db\_xref="GI:5002494"  
 /translation="MRLLTLGLCGSVATPLGPKWPEVFGRLASPGFGEVANDQERRWTLTAAPG  
 RRLVYFTHDELSHLCYDVKLSSGAKVLATCGDSTDERAPGKDFYSLGSS  
 KDTYSLGSLDITRRSDYSNKKPFTGFAFYAAEDIDECQVAPGFAFPCDHCHNHLCGFCSCNAG  
 GGFYVSCRAGYLVYLHRRKTCSEQSL"  
 27..71  
 /gene="sMAP"

sig\_peptide

mat\_peptide 72..581  
/gene="sMAP"  
/product="unnamed"  
polya\_signal 706..711  
/gene="sMAP"  
BASE COUNT 126 a 268 c 208 g 123 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6 83e-16 Length: 725  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-874-198-1 (1-41) x AB008047 (1-725)  
1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
72 ACCCCCTGGGGCGGAAGTGCGCTGAACCTGTCTCGGGCGCTGGCATCCCGGCTTT 131  
|||||  
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProGlyTyr 40  
|||||  
Db 132 CCAGGGAGATGTCATGACACAGAGCGCGCTGACCTGACTGCACCCCGGCTAC 191  
|||||  
QY 41 Arg 41  
|||  
Db 192 CGC 194  
RESULT 3  
HOSA18283 729 bp mRNA linear PRI 12-JUL-1999  
LOCUS  
DEFINITION Homo sapiens mRNA for mannose binding lectin-associated serine  
protease-2, alternatively spliced transcript (clone phi-6).  
VERSION Y18283.1 GI:5459318  
KEYWORDS lectin-associated protein; mannose binding lectin-associated  
protein; mannose binding protein; serine protease.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 729)  
AUTHORS Stover,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T.,  
Jenssens,J.C. and Schwaebler,W.J.  
TWO constituents of the initiation complex of the mannan-binding  
lectin activation pathway of complement are encoded by a single  
structural gene  
J. Immunol. 162 (6), 3481-3490 (1999)  
JOURNAL  
MEDLINE 99192764  
PUBMED 10092804  
REFERENCE 2 (bases 1 to 729)  
AUTHORS Schwaebler,W.J.  
TITLE Direct Submission  
SUBMITTED (19-OCT-1998) W.J. Schwaebler, Department of Microbiology  
and Immunology, University of Leicester, University Road, PO Box  
138, Leicester LE1 9HN, UK  
Related sequence Y09926.  
COMMENT  
FEATURES  
source  
1..729  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p36.2-3"  
/clone="phi-6"  
/tissue\_type="liver"  
1..720  
/gene="MASP-2"  
1..15  
/gene="MASP-2"  
10..573  
/gene="MASP-2"  
CDS  
gene  
5' UTR  
CDS

/note="alternative"  
/codon\_start=1  
/product="mannose binding lectin-associated serine  
protease-2 related protein, Map19 (19kDa)"  
/protein\_id="CA50730.1"  
/db\_xref="GI:5459319"  
/translation="MRLTRLLGLGCVATPLGPKPEVFGRLASPGFPGEVANDOE  
RMTLTPAPGYRLRYFTHTDLESLICEYDFVLTSSGAKVLATLGGSTPTERAPG  
KDFYSGSSLDITFRSDYSNEKPTFGFAFYAEIDECQVAPGAPICDHHCINHL  
GGRYCSKAGYVLHNRKTCSEOSTL  
16..60  
/gene="MASP-2"  
61..570  
/gene="MASP-2"  
/product="mannose binding lectin-associated serine  
protease-2 related protein, Map19 (19kDa)"  
574..720  
/gene="MASP-2"  
633..698  
/gene="MASP-2"  
16..264 c 201 g 122 t  
BASE COUNT 142 a 264 c 201 g 122 t  
ORIGIN  
polya\_signal  
/gene="MASP-2"  
16..698  
Alignment Scores:  
Pred. No.: 6 87e-16 Length: 729  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-874-198-1 (1-41) x HOSA18283 (1-729)  
QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
Db 61 ACCCCCTAGCGCCGGAAGTGCGCTGAACCTGTCTCGGGCGCTGGCATCCCGGCTTT 120  
|||||  
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProGlyTyr 40  
|||||  
Db 121 CCAGGGAGATGTCATGACACAGAGCGCGCTGACCTGACTGCACCCCGGCTAC 180  
|||||  
QY 41 Arg 41  
|||  
Db 181 CGC 183  
RESULT 4  
HOSA18281 736 bp mRNA linear PRI 12-JUL-1999  
LOCUS  
DEFINITION Homo sapiens mRNA for mannose binding lectin-associated serine  
protease-2, alternatively spliced transcript (clone phi-5).  
VERSION Y18281.1 GI:5459314  
KEYWORDS lectin-associated protein; mannose binding lectin-associated  
protein; mannose binding protein; serine protease.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 736)  
AUTHORS Stover,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T.,  
Jenssens,J.C. and Schwaebler,W.J.  
TWO constituents of the initiation complex of the mannan-binding  
lectin activation pathway of complement are encoded by a single  
structural gene  
J. Immunol. 162 (6), 3481-3490 (1999)  
JOURNAL  
MEDLINE 99192764  
PUBMED 10092804  
REFERENCE 2 (bases 1 to 736)  
AUTHORS Schwaebler,W.J.  
TITLE Direct Submission  
SUBMITTED (19-OCT-1998) W.J. Schwaebler, Department of Microbiology  
and Immunology, University of Leicester, University Road, PO Box  
138, Leicester LE1 9HN, UK

REMARK Revised by author (17-JUL-1999)  
COMMENT Related sequence Y09926.  
FEATURES  
Location/Qualifiers  
source  
1..736  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p36.2-3"  
/clone="phl-5"  
/tissue\_type="liver"  
1..736  
/gene="MASP-2"  
1..21  
/gene="MASP-2"  
22..579  
/gene="MASP-2"  
/note="alternative"  
/codon\_start=1  
/product="mannose binding lectin-associated serine  
protease-2 related protein, MAP19 (19kDa)"  
/protein\_id="CAB50728.1"  
/db\_xref="GI:5459315"  
/translation="MKLLFLGLGCVATPLGPKPPEVFGRLASPGFGEYANDOE  
KRMFLAPPGYRLRYFTFHDLESLICEYDVKLSGAKVLATLGGQSTPTERAPG  
KDFYSLGSSLDITFRSDYSNEKPFGEFAFYAAEDIDECQVAPGEAPTCDDHCHNHL  
GGFYCSGRAGYVLRNKRKTCSEOSL"  
581..736  
/gene="MASP-2"  
700..705  
/gene="MASP-2"  
BASE COUNT 144 a 266 c 203 g 123 t  
ORIGIN  
3'UTR  
polyA\_signal  
Alignment Scores:  
Pred. No.: 5.93e-16 Length: 736  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-874-198-1 (1-41) x HOSA18281 (1-736)  
QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
Db 67 ACCCCCTTAGCCCGGAAGTGGCCTGAACCTGTGTGGGGCGCTGGCATCCCCGGCTTT 126  
|||||  
Ov 21 ProGlyGluTyrAlaAsnAspGlnGluArgTyrThrLeuThrAlaProProGlyTyr 40  
|||||  
127 CCAGGGGAGATGCAATGACGAGAGCGGCGCTGACCTGACCTGACCCCGGCTAC 186  
|||||  
41 Arg 41  
|||  
Db 187 CGC 189  
RESULT 5  
LOCUS HOSA18284 1184 bp mRNA linear PRI 12-JUL-1999  
DEFINITION Homo sapiens mRNA for mannose binding lectin-associated serine  
protease-2, incompletely spliced primary transcript, clone phl-8.  
ACCESSION Y18284  
VERSION Y18284.1 GI:5459320  
KEYWORDS lectin-associated protein; mannose binding lectin-associated  
protein; mannose binding protein; serine protease.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Storer,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T.,  
Jensenius,J.C. and Schwaebble,W.J.  
TITLE Two constituents of the initiation complex of the mannan-binding  
lectin activation pathway of complement are encoded by a single

JOURNAL structural gene  
MEDLINE J. Immunol. 162 (6), 3481-3490 (1999)  
PUBMED 99192764  
REFERENCE 10092804  
2 (bases 1 to 1184)  
AUTHORS Schwaebble,W.J.  
TITLE Direct Submission  
JOURNAL Submitted (19-OCT-1998) W.J. Schwaebble, Department of Microbiology  
and, Immunology, University of Leicester, University Road, PO Box  
138, Leicester LE1 9HN, UK  
COMMENT Related sequence Y09926  
FEATURES  
Location/Qualifiers  
source  
1..1184  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p36.2-3"  
/clone="phl-8"  
/tissue\_type="liver"  
1..1184  
/gene="MASP-2"  
1..21  
/gene="MASP-2"  
1024..1184  
/gene="MASP-2"  
1144..1149  
/gene="MASP-2"  
BASE COUNT 201 a 430 c 325 g 228 t  
ORIGIN  
3'UTR  
polyA\_signal  
Alignment Scores:  
Pred. No.: 1.09e-15 Length: 1184  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-874-198-1 (1-41) x HOSA18284 (1-1184)  
QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
Db 67 ACCCCCTTAGCCCGGAAGTGGCCTGAACCTGTGTGGGGCGCTGGCATCCCCGGCTTT 126  
|||||  
Ov 21 ProGlyGluTyrAlaAsnAspGlnGluArgTyrThrLeuThrAlaProProGlyTyr 40  
|||||  
127 CCAGGGGAGATGCAATGACGAGAGCGGCGCTGACCTGACCTGACCCCGGCTAC 186  
|||||  
41 Arg 41  
|||  
Db 187 CGC 189  
RESULT 6  
LOCUS HSMASP2 2450 bp mRNA linear PRI 07-JUL-1998  
HSMASP2



DEFINITION H.sapiens mRNA for mannan-binding lectin-associated serine  
protease-2.  
ACCESSION X98400  
VERSION X98400.1 GI:3297878  
KEYWORDS lectin-associated protein; mannan-binding protein; MASP; serine  
protease.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2450)  
Thiel,S., Vorup-Jensen,T., Stover,C.M., Schwaebler,W., Laursen,S.B.,  
Poulsen,K., Willis,A.C., Eggleton,P., Hansen,S., Holmskov,U.,  
Reid,K.B.M. and Jensenius,J.C.  
Identification and characterization of a novel protein of the human  
complement system, mannan-binding lectin-associated serine  
protease-2 (MASP-2)  
Unpublished  
2 (bases 1 to 2450)  
Thiel,S.  
Direct Submission  
Submitted (08-JUN-1996) S. Thiel, University of Aarhus, Dept. of  
Medical Microbiology, Immunol., Bartholin Building, Wilhelm Meyers  
Alle, 8000 Aarhus C, Denmark  
Location/Qualifiers  
1. 2450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="4"  
37. 2097  
/note="MASP-2"  
/codon\_start=1  
/product="mannan-binding lectin-associated serine  
protease-2"  
/protein\_id="CA67050.1"  
/db\_xref="GI:3297879"  
/db\_xref="SPTREMBL:O75754"  
/translation="MRLLTLGLLGLGVATPLGPKMPPEVFGRLASPGFGEVANDOE  
RRMTLTPAGYRLRLFTTHDELHLCEYDFPVKLSAKYLATLTCGSESTDEAPG  
KDTFYSLSGSLDITFRSDYSNEKPTGEAFVAAEDIDECVAPGEATPCHHCHNLH  
GGFYCSRCAGVYLHNRKRTCSALSGOYFTRSGELSPERPRPKLSQSTYSISLE  
EGFSVILDFVESFDETHPEFLCPYDFLQTDREHGFPGKTLPHRIETKSNVTI  
TEVTDSEGDHGWKIHYSIAHACYPMAPNGHVSYPQAKYILKDSFISCEGYEL  
LOGHLPLKSFPAVOCOKGSDMPMPACSIIDCGPPDLPSGRVEYITGPGVITYKAVI  
OYSCERTFYTKKLVNDGKYVCADGFWTSKGEKSLPVEPCGSLASRTTGGRTYGGOK  
AKPGDFPMQVLLIGGTTAGALLDYNNVLTAAHAYEOKHDSALDITMGTLKRLSPH  
YTQAMSEAVFIHEGYTHDAGFNDIALIKLNKVVYINSNTIPCLPRKESFMTD  
IGTASGWLTORGLFARLNLVVDIPVDHOKCTAAVEKPPYPRGSVTANMLCAGLESG  
GKDCSRGDSGALVFLDSETEFWFVGIVSGMNCBAGOGYGVYTKVINYIPIENI  
ISDF"

BASE COUNT 609 a 644 c 617 g 580 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2.19e-15 Length: 2450  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservation: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x HSMASP2 (1-2450)

OY 1 ThrProLeuGlyProIyTirProGluProValIPheGlyArgLeuAlaSerProGlyPhe 20  
DB 82 ACCCCCTTAGGCGCCAGAGTGCCGACCTGTCTGGCGGCGCTGGCACCCTCCCGGCTTT 141  
OY 21 ProGlyGluTyTAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyT 40  
DB 142 CCAGGGAGATATGCCAATGACAGGAGCGCGCTGACCTGACCTGACACCCCGGCTAC 201  
OY 41 Arg 41  
|||

Db 202 CGC 204

RESULT 7  
LOCUS HSMASP2PR 2455 bp mRNA linear PRI 12-DEC-1998  
DEFINITION Homo sapiens mRNA for MASP-2 protein.  
ACCESSION Y09926  
VERSION Y09926.1 GI:4007626  
KEYWORDS MASP-2 gene.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2455)  
Jensen,T.V.  
Direct Submission  
Submitted (06-DEC-1996) T.V. Jensen, University of Aarhus,  
Microbiology & Immunology, Bartholin Building, 8000 Aarhus C,  
DENMARK  
Revised by [3]  
2 (bases 1 to 2455)  
Thiel,S., Jensen,T.V., Stover,C.M., Schwaebler,W.J., Laursen,S.B.,  
Poulsen,K., Willis,A.C., Eggleton,P., Hansen,S., Holmskov,U.,  
Reid,K.B.M. and Jensenius,J.C.  
A second serine protease associated with mannan-binding lectin that  
activates complement  
Nature 386 (6624), 506-510 (1997)  
9727412  
MEDLINE  
PUBMED 9087411  
REFERENCE 3 (bases 1 to 2455)  
Schwaebler,W.J.  
Direct Submission  
Submitted (08-DEC-1998) W.J. Schwaebler, Department of Immunology  
and Microbiology, University of Leicester, University Road,  
Leicester LE1 9HN, UK  
On Dec 13, 1998 this sequence version replaced gi:1929053.  
Location/Qualifiers  
1. 2455  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p36.2-3"  
/tissue\_type="liver"  
1. 2455  
/gene="MASP-2"  
1. 16  
/gene="MASP-2"  
17. 2077  
/gene="MASP-2"  
/codon\_start=1  
/evidence="experimental"  
/product="MASP-2 protein"  
/protein\_id="CA71059.1"  
/db\_xref="GI:4007627"  
/db\_xref="SPTREMBL:O00187"  
/translation="MRLLTLGLLGLGVATPLGPKMPPEVFGRLASPGFGEVANDOE  
RRMTLTPAGYRLRLFTTHDELHLCEYDFPVKLSAKYLATLTCGSESTDEAPG  
KDTFYSLSGSLDITFRSDYSNEKPTGEAFVAAEDIDECVAPGEATPCHHCHNLH  
GGFYCSRCAGVYLHNRKRTCSALSGOYFTRSGELSPERPRPKLSQSTYSISLE  
EGFSVILDFVESFDETHPEFLCPYDFLQTDREHGFPGKTLPHRIETKSNVTI  
TEVTDSEGDHGWKIHYSIAHACYPMAPNGHVSYPQAKYILKDSFISCEGYEL  
LOGHLPLKSFPAVOCOKGSDMPMPACSIIDCGPPDLPSGRVEYITGPGVITYKAVI  
OYSCERTFYTKKLVNDGKYVCADGFWTSKGEKSLPVEPCGSLASRTTGGRTYGGOK  
AKPGDFPMQVLLIGGTTAGALLDYNNVLTAAHAYEOKHDSALDITMGTLKRLSPH  
YTQAMSEAVFIHEGYTHDAGFNDIALIKLNKVVYINSNTIPCLPRKESFMTD  
IGTASGWLTORGLFARLNLVVDIPVDHOKCTAAVEKPPYPRGSVTANMLCAGLESG  
GKDCSRGDSGALVFLDSETEFWFVGIVSGMNCBAGOGYGVYTKVINYIPIENI  
ISDF"

gene  
5' UTR,  
CDS  
3' UTR

sig\_peptide  
3' UTR

polyA\_signal 2278..2283  
/gene="MASP-2"

BASE COUNT 623 a 639 c 612 g 581 t

ORIGIN

Alignment Scores:

Pred. No.:	2196-15	Length:	2455
Score:	238.00	Matches:	41
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-874-198-1 (1-41) x HSMASP2PR (1-2455)

QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
62 ACCCCTTAGGCCGCAAGAGGCGCTGAACCTGTGTCGGCGCGCTGACATCCCCGGCTTT 121

21 ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrPheLeuThrAlaProProGlyTyr 40  
|||||  
122 CCAGGGAGATATGCCAATACACAGAGCGCGCTGACCTGACCTGACCTGACCTGACCTGAC 181

QY 41 Arg 41  
|||

DB 182 CGC 184

RESULT 8

AB033742 2819 bp DNA linear PRI 14-APR-2000

LOCUS Homo sapiens sMAP, MASP2 genes for small MBL-associated protein,  
DEFINITION MBL-associated serine protease(MASP)-2, complete and partial cds.  
AB033742  
VERSION AB033742.1 GI:6092071  
KEYWORDS MASP2; sMAP; MBL-associated serine protease(MASP)-2; small  
MBL-associated protein.  
SOURCE Homo sapiens male DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2819) Takahashi,M. and Fujita,T.  
AUTHORS Takahashi,M. and Fujita,T.  
TITLE Partial genomic structure of human MBL-associated serine protease  
(MASP)-2 (from exon 1 to exon 5)  
JOURNAL Published Only in Database (1999)  
REFERENCE 2 (bases 1 to 2819)  
AUTHORS Takahashi,M. and Fujita,T.  
TITLE Direct Submission  
AUTHORS Submitted (20-OCT-1999) Minoru Takahashi, Fukushima Medical  
COLLEGE, Department of Biochemistry, 1 Hikarigaoka, Fukushima,  
Fukushima 960-1295, Japan (E-mail:minoru@med.fmu.ac.jp,  
Tel:81-24-548-2111(ex 2232), Fax:81-24-548-6760)  
FEATURES  
source location/Qualifiers  
1..2819  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="male"  
1..2433  
/gene="MASP2"  
/note="sMAP"  
<1..34  
/number="1"  
/gene="MASP2"  
/number=1  
join(30..34,116..344,501..678,1693..1824,2279..2292)  
/gene="MASP2"  
/codon\_start=1  
/product="small MBL-associated protein"  
/protein\_id="BA085658.1"  
/db\_xref="GI:6092072"  
/translation="MRLTLGLGCGSVATPLGPKMPPEVFGRLASPGFPEYANDOE  
RRMLTAPGRLKLYFTHFDLELSHCEYDFVKLSGAKVLAATLCGESDTERAPG  
KDFYSGSSLDITFRSDYSNEKPTGEAFYAEDIDECOVARGEAPTCDHCHNHL  
GGFYCSGRAGYVLRNKRKTCSEOSL"

CDS

join(30..34,116..344,501..678,1693..>1824)  
/gene="MASP2"  
/codon\_start=1  
/product="MBL-associated serine protease(MASP)-2"  
/protein\_id="BA085658.1"  
/db\_xref="GI:6092073"  
/translation="MRLTLGLGCGSVATPLGPKMPPEVFGRLASPGFPEYANDOE  
RRMLTAPGRLKLYFTHFDLELSHCEYDFVKLSGAKVLAATLCGESDTERAPG  
KDFYSGSSLDITFRSDYSNEKPTGEAFYAEDIDECOVARGEAPTCDHCHNHL  
GGFYCSGRAGYVLRNKRKTCSEOSL"

35..115  
/gene="MASP2"  
116..344  
/gene="MASP2"  
/number=2  
345..500  
/gene="MASP2"  
501..678  
/gene="MASP2"  
/number=3  
679..1692  
/gene="MASP2"  
1693..1824  
/gene="MASP2"  
/number=4  
1825..2278  
/gene="MASP2"  
2279..2433  
/gene="MASP2"  
/note="specific exon for sMAP"  
/number=5  
2414..2419  
/gene="MASP2"  
2434..>2819

polyA\_signal

intron  
BASE COUNT 510 a 924 c 830 g 538 t 17 others  
ORIGIN

Alignment Scores:

Pred. No.:	2,56-15	Length:	2819
Score:	238.00	Matches:	41
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-874-198-1 (1-41) x AB033742 (1-2819)

QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
DB 156 ACCCCTTAGGCCGCAAGAGGCGCTGAACCTGTGTCGGCGCGCTGACATCCCCGGCTTT 215

QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrPheLeuThrAlaProProGlyTyr 40  
|||||  
DB 216 CCAGGGAGATATGCCAATACACAGAGCGCGCTGACCTGACCTGACCTGACCTGACCTGAC 275

QY 41 Arg 41  
|||

DB 276 CGC 278

RESULT 9

H05A18286 4464 bp DNA linear PRI 12-JUL-1999

LOCUS H05A18286  
DEFINITION Homo sapiens partial MASP-2 gene, clone pgm-2A.  
ACCESSION Y18286  
VERSION Y18286.1 GI:5459322  
KEYWORDS lectin-associated protein; mannose binding lectin-associated  
protein; mannose binding protein; serine protease.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 4464)  
AUTHORS Stover,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T.,

TITLE Jensenius, J.C. and Schwaebble, W.J.  
Two constituents of the initiation complex of the mannan-binding  
lectin activation pathway of complement are encoded by a single  
structural gene  
JOURNAL Immunol. 162 (6), 3481-3490 (1999)  
MEDLINE 99192764  
PUBMED 10092804  
REFERENCE 2 (bases 1 to 4464)  
AUTHORS Schwaebble, W.J.  
TITLE Direct Submission  
JOURNAL Submitted (18-OCT-1998) W.J. Schwaebble, Department of Microbiology  
and, Immunology, University of Leicester, University Road, PO Box  
138, Leicester LE1 9HN, UK

FEATURES  
source

1..4464  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p36.2-3"  
/clone="pgm-2A"  
/cell\_type="lymphocyte"  
1..4464  
/gene="MASP-2"  
join(1..559,2414..2610,2926..4464)  
/gene="MASP-2"  
/note="alternative"  
join(1..559,1004..1156)  
/gene="MASP-2"  
/note="alternative"  
1..359  
/gene="MASP-2"  
/note="a"  
join(16..559,2414..2610,2926..4245)  
/gene="MASP-2"  
/note="alternative"  
/codon\_start=1  
/product="MASP-2 protein"  
/protein\_id="CAB50733.1"  
/db\_xref="GI:5459324"  
/translation="MRLTLILGLGSSVATPLGPKPEPVFGRLASPGPGEYANDOE  
RMTLTAPPGYRLRLFTFHLDELHSLCEYDVKSSGAKVLATILGQESTDTERAPG  
KTFYSLGSSLDITERSDYSNEKPTFGFAFYAAEDIDECQVAPGAPPCDHNHNL  
GGEYSCRAGYVLRNRKRTCSALCSGVFTORSGLSSPEYPRPYRLSSKYISLE  
EGFSYVLDESEFVETTHPEETLCPYDFLKIQDREHGFEGCKTLPHRIETKSNVTI  
TFTYDSGDHCKMKTHTYSTAHACPYBMAPPNCHVSPVNAKYILKNSFISCTGVEL  
LQCHLPLKSTFANCCQKDSMDRMPRAIVYCCRPDDCSGSGREYITTGCVTTKAVI  
QSCERTFTYMKVNDGKIVCEADGFWTSSKGEKSLPCEPVCGLSARTTGGRTYGGOK  
AKPGDPWQVIGGTTAAGALLILYDMWVLTAAHAYEOKHDASALDIRGTLRSLSPH  
YQWAMEAVFIHEGYTHDAGFNDIALILKNNVYNSITPILCRKEAESPMRTD  
IGTASGMGLTORGLARNLAMYDIPVDHOKCTAAVEKRPYPRGSVTANMLCGLESG  
GDSCHGDSGALVFLDSETERFVGGIYSWMSNGEACQYCVYTKVINYIMIENTI  
ISDF"  
join(16..559,1004..1017)  
/gene="MASP-2"  
/note="alternative"  
/codon\_start=1  
/product="Mab19 protein"  
/protein\_id="CAB50732.1"  
/db\_xref="GI:5459323"  
/translation="MRLTLILGLGSSVATPLGPKPEPVFGRLASPGPGEYANDOE  
RMTLTAPPGYRLRLFTFHLDELHSLCEYDVKSSGAKVLATILGQESTDTERAPG  
KTFYSLGSSLDITERSDYSNEKPTFGFAFYAAEDIDECQVAPGAPPCDHNHNL  
GGEYSCRAGYVLRNRKRTCSGSL"  
1004..1156  
/gene="MASP-2"  
/note="b"  
2414..2610  
/gene="MASP-2"  
/note="c"  
2926..4464  
/gene="MASP-2"  
/note="d"  
938 a 1338 c 1217 g 971 t

## ORIGIN

## Alignment Scores:

Pred. No.:	3,87e-15	Length:	4464
Score:	238.00	Matches:	41
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-874-198-1 (1-41) x HOSA18286 (1-4464)

Qy 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20

Db 61 ACCCCCTTAGCCCGGAGTGGCCCTGACCTGTCTGGCGCCCTGGCATCCCGGCTTT 120

Qy 21 ProGluGlyTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr 40

Db 121 CCAGGCGAGTATGCCCATGACACAGACGCGCTGACCTGACCTGACCTGACCTGACCTGAC 180

Qy 41 Arg 41

Db 181 CGC 183

## RESULT 10

HOSA18287 4465 bp DNA linear PRI 12-JUL-1999

LOCUS HOSA18287

VERSION Y18287.1 GI:5459325

DEFINITION Homo sapiens partial MASP-2 gene, clone pgm-2B.

KEYWORDS lectin-associated protein; mannos binding protein; serine protease.

## SOURCE

## ORGANISM

Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Stover, C.M., Thiel, S., Thelen, M., Lynch, N.J., Vorup-Jensen, T.,

Jensenius, J.C. and Schwaebble, W.J.  
Two constituents of the initiation complex of the mannan-binding  
lectin activation pathway of complement are encoded by a single  
structural gene  
JOURNAL Immunol. 162 (6), 3481-3490 (1999)  
MEDLINE 99192764  
PUBMED 10092804  
REFERENCE 2 (bases 1 to 4465)  
AUTHORS Schwaebble, W.J.  
TITLE Direct Submission  
JOURNAL Submitted (19-OCT-1998) W.J. Schwaebble, Department of Microbiology  
and, Immunology, University of Leicester, University Road, PO Box  
138, Leicester LE1 9HN, UKFEATURES  
source

1..4465  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p36.2-3"  
/clone="pgm-2B"  
/cell\_type="lymphocyte"  
1..4465  
/gene="MASP-2"  
join(1..559,2415..2611,2927..4465)  
/gene="MASP-2"  
join(1..559,1004..1156)  
/gene="MASP-2"  
1..559  
/gene="MASP-2"  
/note="a"  
join(16..559,2415..2611,2927..4246)  
/gene="MASP-2"  
/codon\_start=1  
/product="MASP-2"  
/protein\_id="CAB50735.1"

/db\_xref="GI:5459327"  
/translation="MRLLTLGLGCVATPLGPKMPPEVFGRLASPGFPGEXANDOE  
RRMTLTAPPGYRLRLFTFHFDLELSHCYDFAFKLSSGAKVLATLGGOSTDERAPG  
KDFYSLGSSLDITFRSDYSENEPFGFAFYAEDIDCOVAPGAPPCDDHCHNHL  
GGFYCSGRAGYVLRNRKRTCSALCSQVFTQNGELSSPEYPRYKLSCTYSILE  
EGFSVLIDVESEDTHEPTELCYPDFLQIDREHGPCCGTLDRHETKSNVTYI  
TEVTDSDGTHGWKIHYSSTAHACPYPMAPNGHVPSPVAKYILKSFISFCTGYEL  
LOGHLPLKSFETAVCOKDSWDRMPALVDCGPPDLCSPPGVEYITGGVTTYKAVI  
OYCEETFTMKVNDGKYVCEADGFWTSKGEKSLVCEPVGSLARTGTGGRTYGGOK  
AKGDFPMTQVLGGTTAAGALLITLYDMVYLAAYVEOKHDAEALDIRGCTLRSLSPH  
YTOANSEAVFIHEGTHDAGFDDIALIKLNKVVINSNTIPCLPRKARESPERDD  
IGTASWGLTQRCFLARNLMTYDIPIVDHOKCYAAVEKPPRGSVTAMNLCGLSNG  
KDSGSCDGGAGLAFVLDSETERWFGIVSWGSMNGEAGQYGVYTYKINYLPMIENI  
ISDF"  
join(16. .559,1004. .1017)  
CDS  
/gene="MASP-2"  
/product="Masp-2"  
/codon\_start=1  
/protein\_id="CAB50734.1"  
/db\_xref="GI:5459326"  
/translation="MRLLTLGLGCVATPLGPKMPPEVFGRLASPGFPGEXANDOE  
RRMTLTAPPGYRLRLFTFHFDLELSHCYDFAFKLSSGAKVLATLGGOSTDERAPG  
KDFYSLGSSLDITFRSDYSENEPFGFAFYAEDIDCOVAPGAPPCDDHCHNHL  
GGFYCSGRAGYVLRNRKRTCSQSL"  
1004. .1156  
/gene="MASP-2"  
/note="b"  
exon  
2415. .2611  
/gene="MASP-2"  
/note="c"  
exon  
2927. .4465  
/gene="MASP-2"  
/note="d"  
BASE COUNT 937 a 1338 c 1219 g 971 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.88e-15 Length: 4465  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-874-198-1 (1-41) x HOSA18287 (1-4465)  
QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
61 ACCCCCTTAGCCCCGGAAGTGGCCTGTCTGCGGGCCCTGGCATCCCCGGGCTTT 120  
ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrThrLeuThrAlaProProGlyTyr 40  
|||||  
Db 121 CCAGGGAGATGATGCCAATGACAGAGCGGCGCTGACCTGACATGACCCCGGCTAC 180  
QY 41 Arg 41  
|||  
Db 181 CGC 183  
RESULT 11  
AF321558S1  
LOCUS AF321558S1 4900 bp DNA linear PRI-18-JAN-2001  
DEFINITION Homo.sapiens MBL-associated serine protease 2 (MASP2) gene, exons 1  
through 6; and MBL-associated protein Map19 (MASP2) gene, complete  
cds, alternatively spliced.  
ACCESSION AF321558  
VERSION AF321558.1 GI:12276131  
KEYWORDS 1 of 5  
SEGMENT  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 4900)

AUTHORS Park,D., Kim,B., Baek,K. and Yoon,J.  
TITLE Structure of Human MASP-2 Gene  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4900)  
AUTHORS Park,D., Kim,B., Baek,K. and Yoon,J.  
TITLE Direct Submision  
JOURNAL Submitted (16-NOV-2000) Genetic Engineering, Kyung Hee University,  
1 Seochun-Ri, Kihung-Up, Yongin City, Kyungki-Do 449-701, Korea  
FEATURES  
source  
location/Qualifiers  
1. .4900  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
join(1. .172,254. .482,639. .816,1830. .1961,2407. .2859)  
/gene="MASP2"  
/product="MBL-associated protein Masp19"  
/note="alternatively spliced"  
1. .172  
/gene="MASP2"  
/number=1  
join(168. .172,254. .482,639. .816,1830. .1961,2407. .2420)  
/gene="MASP2"  
/note="alternatively spliced"  
/codon\_start=1  
/product="MBL-associated protein Map19"  
/protein\_id="MAG50275.1"  
/db\_xref="GI:12276137"  
/translation="MRLLTLGLGCVATPLGPKMPPEVFGRLASPGFPGEXANDOE  
RRMTLTAPPGYRLRLFTFHFDLELSHCYDFAFKLSSGAKVLATLGGOSTDERAPG  
KDFYSLGSSLDITFRSDYSENEPFGFAFYAEDIDCOVAPGAPPCDDHCHNHL  
GGFYCSGRAGYVLRNRKRTCSQSL"  
254. .482  
/gene="MASP2"  
/number=2  
639. .816  
/gene="MASP2"  
/number=3  
1830. .1961  
/gene="MASP2"  
/number=4  
2407. .2859  
/gene="MASP2"  
/note="alternatively spliced"  
/number=5  
3818. .4014  
/gene="MASP2"  
/note="alternatively spliced"  
/number=5  
4328. .4474  
/gene="MASP2"  
/number=6  
BASE COUNT 940 a 1604 c 1374 g 982 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4.24e-15 Length: 4900  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-874-198-1 (1-41) x AF321558S1 (1-4900)  
QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
Db 294 ACCCCCTTAGCCCCGGAAGTGGCCTGTCTGCGGGCCCTGGCATCCCCGGGCTTT 353  
ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrThrLeuThrAlaProProGlyTyr 40  
|||||  
Db 354 CCAGGGAGATGATGCCAATGACAGAGCGGCGCTGACCTGACATGACCCCGGCTAC 413  
QY 41 Arg 41  
|||

Db 414 CGC 416

RESULT 12  
LOCUS HSJ635E18/c

DEFINITION Human DNA sequence from clone RP4-635E18 on chromosome 11p36.11-36.31, complete sequence.

ACCESSION AL109811

VERSION AL109811.40 GI:181523762

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 112769)

AUTHORS Wallis, J.

TITLE Direct Submission

JOURNAL Submitted (11-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On Jan 15, 2002 this sequence version replaced gi:11967852. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information from the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C.elegans/wormep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr1 RP4-635E18 is from the library RPL-4 constructed by the group of Plier de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pCYPAC2

RES

source

1. 112769

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="p36.11-36.31"

/clone="RP4-635E18"

/clone\_1ib="RPL-4"

2691. 3057

/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."

misc-feature

3696. 36769

/note="Single clone region. Assembly confirmed by restriction digest data. Single read sequenced with dGTP" 37125..37470

misc-feature

/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."

misc-feature

38084

/note="Forced join. Gap sized to be approximately 339bp from restriction digest data and spanning pUC clone."

misc-feature

56138..56144

/note="sequence from overlapping clone d0576K7

(AL359082). Assembly confirmed by restriction digest data."

BASE COUNT 29406 a 25975 c 28105 g 29283 t

ORIGIN

Alignment Scores:

pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
8,46e-14	238.00	100.00%	100.00%	100.00%	9
Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
112769	41	0	0	0	0

US-09-874-198-1 (1-41) x HSJ635E18 (1-112769)

QY 1 ThrProLecAlyProLystrProGluProValPheGlyArgLeuAlaSerProGlyPhe 20

Db 43427 ACCCCCTTGCGCCGGAAGTGGCTGACCTGTGTGCGCGCCCTGCATCCCGGCTTT 43368

QY 21 ProGlyGlyTyrAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProGlyTyr 40

Db 43367 CCAGGGAGATATGCCAATGACCAAGAGCGCGCTGGACCTGACTGACACCCCGGCTTAC 43308

QY 41 Arg 41

Db 43307 CGC 43305

RESULT 13

LOCUS MMU250369 733 bp mRNA linear ROD 26-JAN-2000

DEFINITION Mus musculus mRNA for mannose binding lectin-associated serine protease-2 related protein, (MASP-2/Mapi9 gene).

ACCESSION AJ250369

VERSION AJ250369.1 GI:6599308

KEYWORDS mannose binding lectin-associated serine protease-2; MASP-2/Mapi9 gene.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 733)

AUTHORS Stover, C.M., Thiel, S., Lynch, N.J. and Schwaebler, W.J.

TITLE The rat and mouse homologues of MASP-2 and Mapi9, components of the lectin activation pathway of complement

JOURNAL J. Immunol. 163 (12), 6848-6859 (1999)

MEDLINE 20054576

PUBMED 10586086

REFERENCE 2 (bases 1 to 733)

AUTHORS Stover, C.M.

TITLE Direct Submission

JOURNAL Submitted (20-OCT-1999) Stover C.M., Dept. of Experimental Immunology, University of Marburg, Deutschhausstrasse 2, 35037 Marburg, GERMANY

COMMENT This mouse Mapi9 mRNA species differs from mRNA in Y19160 in its 5' end.

FEATURES

source

1. 733

/organism="Mus musculus"

/varietal="C57Bl6"

/db\_xref="taxon:10090"

1. 31

/gene="mouse Mapi9"

1..31

/gene="mouse Mapi9"

/function="unknown"

32..589

/gene="MASP-2/Mapi9"

32..589

/gene="MASP-2/Mapi9"

/function="unknown"

/codon\_start=1

/product="mouse Mapi9"

/protein\_id="CAB63701.1"

```
/db_xref="GI:6599309"
/translation="MRLLFLGLMSLVATLLGSKMPEPVFGRLVSPGFPEKADHD
RSMVTLAPPGYRLRYFTTHFDLELSRYCEYDVKLSSGTVALTLGGESTPTQAPG
NRTFYSLSGSLKVTYTHSDYSNKRPTGFEAFVAAEDVDECRVSLGDSVPCDHCHMYL
GGYGCSCRAGYVLHONKHTECSQSL"
sig_peptide
32..76
/gene="MASP-2/Map19"
mat_peptide
77..586
/gene="MASP-2/Map19"
/gene="mouse Map19"
3'UTR
590..733
/gene="mouse Map19"
590..733
/function="unknown"
polya_signal
718..723
/gene="mouse Map19"
BASE COUNT 163 a 226 c 182 g 162 t
ORIGIN
...
Percent Similarity: 183.00 Matches: 733
Best Local Similarity: 85.37% Conservative: 3
Query Match: 76.05% Mismatches: 6
Indels: 0
Gaps: 0
DB: 10
US-09-874-198-1 (1-41) x MMU250369 (1-733)
QY 1 ThProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
DB 77 ACACCTCTGGGTTCAAAGGCGCTGAACCGTATTCGGGCGCTGGTCCCTGGCTTC 136
QY 21 ProGlyGluTyrAlaAsnSpgIngluArgArgTrpThleuThralaProGlyTyr 40
DB 137 CCAGAGAAGTATCTGACCATCAAGATCGATCCTGACACTGACCTGACCCCTGGCTAC 196
QY 41 Arg 41
DB 197 CGC 199
RESULT 14
LOCUS MMU19160 817 bp mRNA linear ROD 31-JAN-2000
DEFINITION Mus musculus mRNA for mannose binding lectin-associated serine
protease-2 related protein, MAP19 (19kDa).
ACCESSION Y19160
VERSION Y19160.1 GI:6688726
KEYWORDS mannose binding lectin-associated serine protease-2; MAP19 gene.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 817)
Stover,C.M., Thiel,S., Lynch,N.J. and Schwaeble,W.J.
The rat and mouse homologues of MASP-2 and Map19, components of the
lectin activation pathway of complement
J. Immunol. 163 (12), 6848-6859 (1999)
JOURNAL
MEDLINE 20054576
PUBMED 10586086
REFERENCE 2 (bases 1 to 817)
Schwaeble,W.J.
Direct Submission
AUTHORS Submitted (13-JUL-1999) W.J. Schwaeble, Department of Immunology,
University of Leicester, University Road,, Leicester LE1 9HN, UK
JOURNAL
FEATURES
source
1..817
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
1..817
/gene="Map19"
```

```
5'UTR 1..74
/gene="Map19"
CDS 75..647
/gene="Map19"
/codon_start=1
/product="mannose binding lectin-associated serine
protease-2"
/protein_id="CAB65247.1"
/db_xref="GI:6688727"
/translation="MSLPQQLILFLGLMSLVATLLGSKMPEPVFGRLVSPGFPEK
ADHQRSMVTLAPPGYRLRYFTTHFDLELSRYCEYDVKLSSGTVALTLGGESTPT
EQAPNDTFYSLGSLKVTYTHSDYSNKRPTGFEAFVAAEDVDECRVSLGDSVPCDH
CHNYLGYYCSCRAGYVLHONKHTECSQSL"
75..134
/gene="Map19"
648..817
/gene="Map19"
773..777
/gene="Map19"
polya_site
BASE COUNT 194 a 248 c 197 g 178 t
ORIGIN
...
Alignment Scores:
Pred. No.: 3.3e-10 Length: 817
Score: 183.00 Matches: 32
Percent Similarity: 85.37% Conservative: 3
Best Local Similarity: 78.05% Mismatches: 6
Query Match: 76.89% Indels: 0
Gaps: 0
DB: 10
US-09-874-198-1 (1-41) x MMU19160 (1-817)
QY 1 ThProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
DB 135 ACACCTCTGGGTTCAAAGGCGCTGAACCGTATTCGGGCGCTGGTCCCTGGCTTC 194
QY 21 ProGlyGluTyrAlaAsnSpgIngluArgArgTrpThleuThralaProGlyTyr 40
DB 195 CCAGAGAAGTATCTGACCATCAAGATCGATCCTGACACTGACCTGACCCCTGGCTAC 254
QY 41 Arg 41
DB 255 CGC 257
RESULT 15
LOCUS MMU19163 1145 bp mRNA linear ROD 16-FEB-2000
DEFINITION Mus musculus partial mRNA for mannose binding lectin-associated
serine protease-2 (MASP-2 gene).
ACCESSION Y19163
VERSION Y19163.1 GI:6688732
KEYWORDS mannose binding lectin-associated serine protease-2; MASP-2 gene.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1145)
Stover,C.M., Thiel,S., Lynch,N.J. and Schwaeble,W.J.
The rat and mouse homologues of MASP-2 and Map19, components of the
lectin activation pathway of complement
J. Immunol. 163 (12), 6848-6859 (1999)
JOURNAL
MEDLINE 20054576
PUBMED 10586086
REFERENCE 2 (bases 1 to 1145)
Schwaeble,W.J.
Direct Submission
AUTHORS Submitted (13-JUL-1999) W.J. Schwaeble, Department of Immunology,
University of Leicester, University Road,, Leicester LE1 9HN, UK
JOURNAL
FEATURES
source
1..1145
Location/Qualifiers
/organism="Mus musculus"
/strain="B6CBAP1/J"
/db_xref="taxon:10090"
```



Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: amadan@systemsbiology.org  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRM Plate: 18 Row: n Column: 9  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6754641.  
Location/Qualifiers

## FEATURES

## CDS

```
1..3070
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:13718 IMAGE:4211293"
/tissue_type="Liver, normal. 5 month old male mouse."
/clone_lib="MCL_CGAP_L19"
/lab_host="DH10B"
/Note="Vector: PCMV-SPORT6"
17..2074
/codon_start=1
/product="Similar to mannan-binding lectin serine protease 2"
/protein_id="AAH13893.1"
/db_xref="GI:15530225"
/translocation="MRL1FLGLMSLAVTLTLLSKMPEVEGRVSPGPEKADHD
RSMVLPAGVRLRLFTFHPDLTSLRCEYDFKLSGGKVALATCGOSTPEOAG
NMTFVLSGSLVETPSDYSNKRPEAFAPAEVDECRSLGSGVCDHCHMYL
GGTYCCRAGYVILHOKHHCALCSQVFTGNSGYSSEPIQPIPKLSSCTYTRLE
DEFSVILDEVSTFHPDAQCPSYSLKIQDNGEHPGCGTLPPIETSHKVTI
FEATDMSGNHTGKIHVYSTARCPDTPAPNPGSISIPVOAIVLXKRFSEVFKTEL
LOGSVLKSFTAVACOKDSWDRMPCSLIIDCGPDLNGHVDYITGGEVTKYKVI
OXSCEETPYTMSNGKYVCEADGFTSSGKRPVCEPCEGISTHTIGRIVGGGPA
KGDPEVOVLIGOTTAAGALIHDMVLTAAVAVERKMASSINIRMGILKRISPH
YTOARPEELFIHGCTHAGAFDNDIALILKKNVTINGSIMPYCLERKASIMETDF
TGVAVAGWGLTQKGLARNLMEVDIPIADHOKCTAVYERKLPGRVSANMLCGLETGG
KDCSRGDSGALVFLDNETQRMFVGIVSGWSINCGADQGYVTKVINYIPMIENI
ISNF"
```

BASE COUNT 836 a 756 c 708 g 770 t

## ALIGNMENT SCORES:

Alignment Scores:  
No.: 1.17e-09 Length: 3070  
Local Similarity: 183.00 Matches: 32  
Conservative: 85.37%  
Best Local Similarity: 76.05% Mismatches: 6  
Query Match: 76.89% Indels: 0  
DB: 10 Gaps: 0

US-09-874-198-1 (1-41) x BC013893 (1-3070)

```
QY 1 ThProleuglyProlystrProgluProvalPheglyArgleuAlaserProglyPhe 20
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 ACACTTCTGGGTTCAAGTGGCTGAACCTGTATCGGGCGCTGCTCCCTGGGCTTC 121

QY 21 ProgluTyraLaasnaSpGInguArgArGTPrThleuThralaProProglyTyr 40
||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 CCAGAGAGATATGCTGACATCAAGATCGATCCTGGACACTGACTGACCCCTGGCTAC 181

QY 41 Arg 41
|||
Db 182 CGC 184
```

RESULT 18 AL591032 205573 bp DNA linear ROD 14-MAR-2002  
LOCUS LQCUS  
DEFINITION Mouse DNA sequence from clone Rp22-211A10 on chromosome 4, complete

ACCESSION AL591032  
VERSION AL591032.19 GI:19571900  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 205573)

AUTHORS Blakey S.  
TITLE Direct Submission  
JOURNAL Submitted (14-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

## COMMENT

On Mar 21, 2002 this sequence version replaced gi:18476682.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em:, EMBL; Sw:,  
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Rp22-211A10 is  
from the RPK2-22 mouse PAC library  
constructed by the group of Pieter de Jong.  
For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6  
This sequence is the entire insert of clone Rp22-211A10.

## FEATURES

## source

Location/Qualifiers  
1..205573  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="Rp22-211A10"  
/clone\_lib="RPCT-22"  
180212  
misc\_feature  
/note="Tandem repeat. Forced join. Gap size estimated to  
be approximately 160bp by restriction digest data."

BASE COUNT 53187 a 50972 c 49570 g 51844 t

## ALIGNMENT SCORES:

Alignment Scores:  
Pred. No.: 6.46e-08 Length: 205573  
Score: 183.00 Matches: 32  
Percent Similarity: 85.37% Conservative: 78.05%  
Best Local Similarity: 76.89% Mismatches: 6  
Query Match: 76.89% Indels: 0  
DB: 10 Gaps: 0

US-09-874-198-1 (1-41) x AL591032 (1-205573)

```
QY 1 ThProleuglyProlystrProgluProvalPheglyArgleuAlaserProglyPhe 20
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89815 ACACTTCTGGGTTCAAGTGGCTGAACCTGTATTCGGGGCGCTGCTCCCTGGGCTTC 89874

QY 21 ProgluTyraLaasnaSpGInguArgArGTPrThleuThralaProProglyTyr 40
||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89875 CCAGAGAGATATGCTGACATCAAGATCGATCCTGGACACTGACTGACCCCTGGCTAC 89934
```

QY 41 Arg 41  
|||  
Db 89935 CGC 89937



DB:	2	gaps:	0
US-09-874-198-1 (1-41) x	AL506969 (1-224312)		
QY	1	ThrpLeugliProI,ystrpProgluProvalPhegiAryAlaAlaserProgiyPhe 20	
Db	22690	ACACTTCGGGTCAAAATGGCCCTGAACCTGATTGCGGGCCCTGTGTCCTCCCTGGCTTC 22749	
QY	21	ProgiygluTyralaAsnaSpGlnIuArgargrTrpThrLeuThrAlaProProgiTytr 40	
Db	22750	CCAGAGAGTATGCTGACCAACATGATCGATTCGTGACACTGACACCCCTGGCTAC 22809	
QY	41	Arg 41	
Db	22810	CGC 22812	
RESULT 20			
LOCUS	RNO277747	2037 bp	mRNA linear ROD 05-Oct-2000
DEFINITION	Rattus norvegicus partial mRNA for mannose-binding protein associated serine protease-2 (masp-2 gene).		
ACCESSION	AJ277747		
VERSION	AJ277747.1	GI:7799288	
KEYWORDS	mannose-binding protein associated serine protease-2; MASP-2 gene.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 2037)		
AUTHORS	Wallis, R. and Dodd, R.B.		
TITLE	Interaction of mannose-binding protein with associated serine proteases: effects of naturally occurring mutations		
JOURNAL	J. Biol. Chem. 275 (40), 30962-30969 (2000)		
MEDLINE	20469449		
PUBMED	10913141		
REFERENCE	2 (bases 1 to 2037)		
AUTHORS	Wallis, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-MAY-2000) Wallis R., Department of Biochemistry, Glycobiology Institute, University of Oxford, South Parks Road, Oxford OX1 3OU, UNITED KINGDOM		
FEATURES	Location/Qualifiers		
source	1..2037		
gene	/organism="Rattus norvegicus"		
CDS	/db_xref="taxon:10116"		
	/tssue_1ib="liver"		
	1..2037		
	/gene="masp-2"		
	<1..2037		
	/gene="masp-2"		
	/function="complement activation"		
	/codon_start=1		
	/product="mannose-binding protein associated serine protease-2"		
	/protein_id="CAB90832.1"		
	/db_xref="GI:7799289"		
	/translation="GILMSLVATLLGSKMPVVRGRVSPGPRKYGNDHDSWTLTMAPRGRRLRFTFHEMLSTKREYTFYVLTSTKVLATLQGESTDTERAPNDTFYSIGPSLAKVFHSDYSNEKPTGTEARVVAEDDECTSLGDSVPCDHYCHNVLGYGCSGRVGLIHQNHKHTSCALSCGVFTYRSGFELSDSPPEYPOPYPLSCAVNRLIEGFSITLDVPSGFVEMHPEAOCEPDSLKIQTDRREYGFPGKTLPRLEIDSNKVTITPTDSESGNHRGKLIHTYSTAOCPDDPTAPPNGHSLPVQATVYLKDSRVSFCKGFELGSLVPLKSPFAVCKOKDSMDRPIPEGLIDCGRPDDLPHNHYITIGPEYTYKAVIYQSCETPTYMSNGKTYCEADGFTWSKGRSLPVCAPVCGLSITHSGRH1IGQDPKPEDFTWQVLLLGTTAAGALIHDDWVLTAAHAYGKTEASSLDIRNGILRLSPHYTQAMPFAVHEIGYTGAGFNDIALIKLKNNKVINRIMPICPRKEAASLAKMTDFGVAGWGVTFOGFLARNLMFEVDIPIVDHOKCATVYTKQPYGAKVYVMMLCAGLDAGKDCRCDSGGGLVLEIDNETQWFMFGGISVMSINCSGSEQGYVYTKVTNYIPIENIINN"		
	<1..36		
	/gene="masp-2"		
	37..2034		
	/gene="masp-2"		
sig_peptide			
mat_peptide			



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 11, 2003, 12:10:24 : Search time 83 Seconds  
(without alignments)  
217.400 Million cell updates/sec

Title: US-09-874-198-1  
Perfect score: 238  
Sequence: 1 TPLGPKMPEPVFGRLASPGF.....GEYANDQERMTLTAPGPR 41

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Search: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09874198/runat\_10012003\_092045\_2707/app.query.fasta\_1.199  
-DB=published\_applications\_NA -OFMT=fastlap -SUFFIX=p2n.tmpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=60 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100  
-THR MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09874198 @CGN 1.1.36 @runat\_10012003\_092045\_2707  
-NCPU=6 -ICPU=3 -NO\_XLPHY -NO\_MMAPP -LARGEOUT=1 -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMECUT=120 -WARN\_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:\*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/FCI_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCUTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	238	100.0	2475	10	US-09-874-198-3
2	238	100.0	2475	10	US-09-874-238-3
3	93.5	39.3	810	10	US-09-764-853-143
4	91.5	38.4	742	10	US-09-833-381-1924

5	91.5	38.4	1988	9	US-10-004-551-21	Sequence 21, Appl
6	91.5	38.4	2127	9	US-10-004-551-17	Sequence 17, Appl
7	91.5	38.4	2127	9	US-10-004-551-19	Sequence 19, Appl
8	91.5	38.4	2143	9	US-10-004-551-23	Sequence 23, Appl
9	91.5	38.4	3863	9	US-10-004-551-13	Sequence 13, Appl
10	91.5	38.4	3879	9	US-10-004-551-15	Sequence 15, Appl
11	83.5	35.1	4344	10	US-09-917-800A-1712	Sequence 1712, Ap
12	81	34.0	1148	10	US-09-925-302-168	Sequence 168, Ap
13	81	34.0	1480	10	US-09-918-447-39	Sequence 39, Appl
14	81	34.0	1480	10	US-09-880-107-2196	Sequence 2196, Ap
15	81	34.0	2386	9	US-09-808-602-92	Sequence 92, Appl
16	81	34.0	2493	10	US-09-880-107-2256	Sequence 2256, Ap
17	81	34.0	2555	12	US-10-044-090-613	Sequence 613, App
18	80	33.6	354	10	US-09-960-352-3224	Sequence 3224, Ap
19	80	33.6	354	10	US-09-960-352-8661	Sequence 8661, Ap
20	80	33.6	425	10	US-09-960-352-10119	Sequence 10119, A
21	80	33.6	433	10	US-09-960-352-14174	Sequence 14174, A
22	79.5	33.4	416	10	US-09-960-352-1998	Sequence 1998, Ap
23	78.5	33.0	344	10	US-09-960-352-14817	Sequence 14817, A
24	78	32.8	390	10	US-09-960-352-5362	Sequence 5362, Ap
25	78	32.8	456	10	US-09-833-381-1925	Sequence 1925, Ap
26	78	32.8	559	10	US-09-833-381-1926	Sequence 1926, Ap
27	78	32.8	1542	10	US-09-822-849A-273	Sequence 273, App
28	78	32.8	1580	10	US-09-925-301-440	Sequence 440, App
29	78	32.8	1992	10	US-09-729-674-91	Sequence 91, Appl
30	76	31.9	18385	10	US-09-764-860-1018	Sequence 1018, Ap
31	74.5	31.3	691	10	US-09-925-300-476	Sequence 476, App
32	74.5	31.3	2647	10	US-09-880-107-2105	Sequence 2105, Ap
33	74.5	31.3	2659	12	US-10-044-090-472	Sequence 472, App
34	74.5	31.3	2725	10	US-09-925-301-182	Sequence 182, App
35	74	31.1	1547	10	US-09-917-800A-1668	Sequence 1668, App
36	70.5	29.6	414	10	US-09-960-352-4891	Sequence 4891, Ap
37	70	29.4	3955	10	US-09-855-722-4	Sequence 4, Appl1
38	70	29.4	4702	10	US-09-880-107-1545	Sequence 1545, Ap
39	69	29.0	2457	10	US-09-850-048A-1	Sequence 1, Appl1
40	69	29.0	3737	10	US-09-747-371-1	Sequence 1, Appl1
41	68	28.6	828	9	US-09-732-242-1	Sequence 1, Appl1
42	68	28.6	1366	10	US-09-800-729-15	Sequence 15, Appl
43	68	28.6	1440	12	US-10-067-422-6	Sequence 20, Appl
44	68	28.6	1635	9	US-10-098-841-207	Sequence 6, Appl1
45	68	28.6	1637	9	US-09-808-602-18	Sequence 207, App
46	68	28.6	1691	9	US-09-808-602-18	Sequence 16, Appl
47	68	28.6	3318	9	US-09-808-602-91	Sequence 91, Appl
48	68	28.6	3546	10	US-09-850-048A-3	Sequence 3, Appl1
49	68	28.6	3919	10	US-09-942-366-4	Sequence 4, Appl1
50	68	28.6	5021	10	US-09-285-385C-1	Sequence 1, Appl1
51	67.5	28.4	2308	10	US-09-917-800A-1348	Sequence 1348, Ap
52	67.5	28.4	4661	10	US-09-285-385C-3	Sequence 3, Appl1
53	67	28.2	4771	10	US-09-942-366-2	Sequence 2, Appl1
54	67	28.2	684	10	US-09-833-381-1923	Sequence 1923, Ap
55	66	27.7	1127	10	US-09-764-853-132	Sequence 132, App
56	65.5	27.5	231	10	US-09-964-824A-775	Sequence 275, Appl
57	65.5	27.5	1806	10	US-09-800-729-75	Sequence 75, Appl
58	65.5	27.5	1806	12	US-10-067-422-5	Sequence 5, Appl1
59	65.5	27.5	2026	9	US-09-905-291A-103	Sequence 103, App
60	65.5	27.5				

## ALIGNMENTS

RESULT 1  
US-09-874-198-3  
Sequence 3, Application US/09874198  
Patent No. US20020082208A1  
GENERAL INFORMATION:  
APPLICANT: Jenseius, Jens Chr.  
APPLICANT: Thiel, Steffen  
TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND  
TITLE OF INVENTION: USES FOR IT  
FILE REFERENCE: 09011-002002  
CURRENT APPLICATION NUMBER: US/09/874,198  
CURRENT FILING DATE: 2001-06-04  
PRIOR APPLICATION NUMBER: 09/054,218

FEATURE:

LOCATION: (1)...(742)  
OTHER INFORMATION: n - A,T,C or G  
US-09-833-381-1924

## Alignment Scores:

Pred. No.: 0.00434 Length: 742  
Score: 91.50 Matches: 21  
Percent Similarity: 46.15% Conservative: 3  
Best Local Similarity: 40.38% Mismatches: 13  
Query Match: 38.45% Indels: 15  
DB: 10 Gaps: 2

US-09-874-198-1 (1-41) x US-09-833-381-1924 (1-742)

QY 5 ProlystP-----ProgluProval----- 11

DB 606 CCCATCTGGGATTCGAAGAGCCGCTGTCATCGCTTGGCGGAGTGATCCGCAAT 547

12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLeuTyraAlaAsnAspGlnGlu 29

546 GCCACACCGCGCGCATCGTCTCCAGGCTCCCGGCAACTACACACCAACCTCACC 487

QY 30 ArgArgTrpThrLeuThrAlaProProGlyTyraArg 41

DB 486 TGTCACTGGCTGCTTGAGGCTCCTCGAGGCCACGCG 451

## RESULT 5

US-10-004-551-21

Sequence 21, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 21

LENGTH: 1988

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(1752)

004-551-21

Alignment Scores:

Pred. No.: 0.0136 Length: 1988  
Score: 91.50 Matches: 21  
Percent Similarity: 46.15% Conservative: 3  
Best Local Similarity: 40.38% Mismatches: 13  
Query Match: 38.45% Indels: 15  
DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-21 (1-1988)

QY 5 ProlystP-----ProgluProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGAGCCGCTGTCATCGCTTGGCGGAGTGATCCGCAAT 1443

QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLeuTyraAlaAsnAspGlnGlu 29

DB 1444 GGCACACCGCGCGCATCGTCTCCAGGCTCCCGGCAACTACACCAACCTCACC 1503

QY 30 ArgArgTrpThrLeuThrAlaProProGlyTyraArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTCGAGGCCACGCG 1539

## RESULT 6

US-10-004-551-17

Sequence 17, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 17

LENGTH: 2127

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(1752)

OTHER INFORMATION: n 2077 can be A, G, C, or T.

US-10-004-551-17

## Alignment Scores:

Pred. No.: 0.0147 Length: 2127  
Score: 91.50 Matches: 21  
Percent Similarity: 46.15% Conservative: 3  
Best Local Similarity: 40.38% Mismatches: 13  
Query Match: 38.45% Indels: 15  
DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-17 (1-2127)

QY 5 ProlystP-----ProgluProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGAGCCGCTGTCATCGCTTGGCGGAGTGATCCGCAAT 1443

QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLeuTyraAlaAsnAspGlnGlu 29

DB 1444 GGCACACCGCGCGCATCGTCTCCAGGCTCCCGGCAACTACACCAACCTCACC 1503

QY 30 ArgArgTrpThrLeuThrAlaProProGlyTyraArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTCGAGGCCACGCG 1539

## RESULT 7

US-10-004-551-19

Sequence 19, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 19

LENGTH: 2127

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(1752)

OTHER INFORMATION: n 2077 can be A, G, C, or T.

US-10-004-551-19

## Alignment Scores:

Pred. No.: 0.0147 Length: 2127

Score: 91.50 Matches: 21  
Percent Similarity: 46.15% Conservative: 3  
Best Local Similarity: 40.38% Mismatches: 13  
Query Match: 38.45% Indels: 15  
DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-19 (1-2127)

OY 5 ProlystP-----ProgiuProval----- 11

DB 1384 CCCTTCGGGATTCAGAGGAGCCGCTGTCATGCTGCGCGGAGTGATCCGCAAT 1443

OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29

DB 1444 GGCACACACCGCGCGCATGCTCTCCAGGCTCCCGGCAACTACAGCAACACCTCACC 1503

OY 30 ArgArgTyrPheLeuThrAlaProProGlyTyrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 1539

US-10-004-551-23

Sequence 23, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

PRIOR FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 23

LENGTH: 2143

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(1755)

OTHER INFORMATION: n 2093 can be A, G, C, or T.

US-10-004-551-23

Alignment Scores:

Pred. No.: 0.0149 Length: 2143

Score: 91.50 Matches: 21

Percent Similarity: 46.15% Conservative: 3

Best Local Similarity: 40.38% Mismatches: 13

Query Match: 38.45% Indels: 15

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 13

LENGTH: 3863

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(3156)

OTHER INFORMATION: n 1755 can be A, G, C, or T

US-10-004-551-13

Alignment Scores:

Pred. No.: 0.0296 Length: 3863

Score: 91.50 Matches: 21

Percent Similarity: 46.15% Conservative: 3

Best Local Similarity: 40.38% Mismatches: 13

Query Match: 38.45% Indels: 15

DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-13 (1-3863)

OY 5 ProlystP-----ProgiuProval----- 11

DB 1384 CCCTTCGGGATTCAGAGGAGCCGCTGTCATGCTGCGCGGAGTGATCCGCAAT 1443

OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29

DB 1444 GGCACACACCGCGCGCATGCTCTCCAGGCTCCCGGCAACTACAGCAACACCTCACC 1503

OY 30 ArgArgTyrPheLeuThrAlaProProGlyTyrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 1539

US-10-004-551-15

Sequence 15, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 15

LENGTH: 3879

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(3159)

OTHER INFORMATION: n 1755 can be A, G, C, or T.

US-10-004-551-15

Alignment Scores:

Pred. No.: 0.0296 Length: 3879

Score: 91.50 Matches: 21

Percent Similarity: 46.15% Conservative: 3

Best Local Similarity: 40.38% Mismatches: 13

Query Match: 38.45% Indels: 15

DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-15 (1-3879)

QY 5 ProlystP-----ProgluProva]----- 11

Db 1384 CCCACTGGGATTCAAGAGCCCGATGATGCCTGCTGGCGGAGTGAATCCGCAT 1443

QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyAlaAsnAspGlnGlu 29

Db 1444 GGCACACCGCGCGCATCTCTCTCCAGGCTTCCCGGCACTACAGCAACACTCACC 1503

QY 30 ArgArgTPrPThLeuThAlaProProGlyTyArg 41

Db 1504 TGTCACTGGCTGCTTGAGCTCCTCGAGCCGACGCG 1539

RESULT 11

US-09-917-800A-1712

Sequence 1712, Application US/09917800A

PATENT No. US20020119462A1

GENERAL INFORMATION:

APPLICANT: Mendrick, Donna

APPLICANT: Porter, Mark

APPLICANT: Johnson, Kory

APPLICANT: Castle, Arthur

APPLICANT: Elashoff, Michael

APPLICANT: Gene Logic, Inc.

TITLE OF INVENTION: Molecular Toxicology Modeling

FILE REFERENCE: 44921-5038-US

CURRENT APPLICATION NUMBER: US/09/917.800A

PRIOR FILING DATE: 2001-07-31

PRIOR APPLICATION NUMBER: US 60/222,040

PRIOR FILING DATE: 2000-07-31

PRIOR APPLICATION NUMBER: US 60/222,880

PRIOR FILING DATE: 2000-11-02

PRIOR APPLICATION NUMBER: US 60/290,029

PRIOR FILING DATE: 2001-05-11

PRIOR APPLICATION NUMBER: US 60/290,645

PRIOR FILING DATE: 2001-05-15

PRIOR APPLICATION NUMBER: US 60/292,336

PRIOR FILING DATE: 2001-05-22

PRIOR APPLICATION NUMBER: US 60/295,798

PRIOR FILING DATE: 2001-06-06

PRIOR APPLICATION NUMBER: US 60/297,457

PRIOR FILING DATE: 2001-06-13

PRIOR APPLICATION NUMBER: US 60/298,884

PRIOR FILING DATE: 2001-06-19

PRIOR APPLICATION NUMBER: US 60/303,459

DR FILING DATE: 2001-07-09

NUMBER OF SEQ ID NOS: 1740

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1712

LENGTH: 4344

TYPE: DNA

ORGANISM: Rattus norvegicus

FEATURE:

OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM\_022849

US-09-917-800A-1712

Alignment Scores:

Pred. No.: 0.331 Length: 4344

Score: 83.50 Matches: 16

Percent Similarity: 41.18% Conservative: 5

Best Local Similarity: 31.37% Mismatches: 17

Query Match: 35.08% Indels: 13

DB: 10 Gaps: 1

US-09-874-198-1 (1-41) x US-09-917-800A-1712 (1-4344)

QY 4 GlyProLysTrpProGluProAlaPhe----- 12

Db 2566 GGTCCAGTTTGGACAGTCTCTTTGTAACATACTTGTGAGGCTTCTCTGACTGGA 2625

QY 13 -----GlyArgLeuAlaSerProGlyPheProGlyGluTyAlaAsnAspGlnGluArg 30

Db 2626 CTCTCTGGGCAATTTTCTAGCCCATCTACTACCCGAGAGCTATCTATATGCGCATGT 2685

QY 31 ArgTPrPThLeuThAlaProProGlyTyArg 41

Db 2686 TTGTGGACATGTGAAGTCCCAAAACAACACTACCG 2718

RESULT 12

US-09-925-302-168

Sequence 168, Application US/09925302

PATENT No. US20020044941A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antbodies

FILE REFERENCE: PA104

CURRENT APPLICATION NUMBER: US/09/925,302

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05918

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1998-03-12

NUMBER OF SEQ ID NOS: 896

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 168

LENGTH: 1148

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (1076)

OTHER INFORMATION: n equals a,t,g, or c

US-09-925-302-168

Alignment Scores:

Pred. No.: 0.144 Length: 1148

Score: 81.00 Matches: 15

Percent Similarity: 58.82% Conservative: 5

Best Local Similarity: 44.12% Mismatches: 14

Query Match: 34.03% Indels: 0

DB: 10 Gaps: 0

US-09-874-198-1 (1-41) x US-09-925-302-168 (1-1148)

QY 8 ProgluProAlaPheGlyArgLeuAlaSerProGlyPheProGlyGluTyAlaAsnAsp 27

Db 139 CCTCAGAGTATTATGGGAGGTGACTTCCCTCTTCCCAAGCTTACCCCAACAC 198

QY 28 GlnGluArgArgTPrPThLeuThAlaProProGlyTyArg 41

Db 199 TTGGAACAACACACTGTGATCAGTCCCGGAGTACAG 240

RESULT 13

US-09-919-497-39

Sequence 39, Application US/09919497

PATENT No. US2002010662A1

GENERAL INFORMATION:

APPLICANT: Mutter, George L.

TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER

FILE REFERENCE: B0801/7225

CURRENT APPLICATION NUMBER: US/09/919,497

CURRENT FILING DATE: 2001-07-31

PRIOR APPLICATION NUMBER: US 60/221,735

PRIOR FILING DATE: 2000-07-31

NUMBER OF SEQ ID NOS: 100

SOFTWARE: PatentIn version 3.0

SEQ ID NO 39

LENGTH: 1480

TYPE: DNA

ORGANISM: Homo sapiens

US-09-919-497-39

Alignment Scores:

Pred. No.: 0.194 Length: 1480

Score: 81.00 Matches: 17  
Percent Similarity: 48.89% Conservative: 5  
Best Local Similarity: 37.78% Mismatches: 13  
Query Match: 34.03% Indels: 10  
DB: 10 Gaps: 1

US-09-874-198-1 (1-41) x US-09-919-497-39 (1-1480)

QY 5 PolystripProgluProvalPhe-----glyarg 14

DB 142 CCCAACTACACGAGACCCGTTCTTCGTCGCGAGGGGATGTGAAGGGGGAATCAGGTTAC 201

QY 15 LeuAlaSerProgluPheProgluTyrrAlaAsnAspGlnuArgArgrTrpThrLeu 34

DB 202 GTGGCAAGTGAAGGGGTCCCAACTCTTACCCCTTAATAGAGAGTCATCTGGACCATA 261

QY 35 ThrAlaProProglu 39

DB 262 ACGGTCCCGAGGGC 276

14 880-107-2196

Sequence 2196, Application US/09880107

Patent No. US20020142981A1

GENERAL INFORMATION:

APPLICANT: Horne, Darci T.

APPLICANT: Vockley, Joseph G.

APPLICANT: Scherf, Uwe

APPLICANT: Gene Logic, Inc.

TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

FILE REFERENCE: 44921-5028-WO

CURRENT APPLICATION NUMBER: US/09/880,107

PRIOR FILING DATE: 2001-06-14

PRIOR APPLICATION NUMBER: US 60/211,379

PRIOR FILING DATE: 2000-06-14

PRIOR APPLICATION NUMBER: US 60/237,054

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 3950

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2196

LENGTH: 1480

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Genbank Accession No. US20020142981A1 E33799

US-09-880-107-2196

Alignment Scores:

No.: 0.194 Length: 1480

Similarity: 48.89% Matches: 17

Local Similarity: 37.78% Conservative: 5

Query Match: 34.03% Mismatches: 13

Indels: 10

Gaps: 1

US-09-874-198-1 (1-41) x US-09-880-107-2196 (1-1480)

QY 5 PolystripProgluProvalPhe-----glyarg 14

DB 142 CCCAACTACACGAGACCCGTTCTTCGTCGCGAGGGGATGTGAAGGGGGAATCAGGTTAC 201

QY 15 LeuAlaSerProgluPheProgluTyrrAlaAsnAspGlnuArgArgrTrpThrLeu 34

DB 202 GTGGCAAGTGAAGGGGTCCCAACTCTTACCCCTTAATAGAGAGTCATCTGGACCATA 261

QY 35 ThrAlaProProglu 39

DB 262 ACGGTCCCGAGGGC 276

RESULT 15

US-09-808-602-92

Sequence 92, Application US/09808602

Patent No. US20020151151A1

GENERAL INFORMATION:

APPLICANT: Vernet, Corine A

APPLICANT: Fernandes, Elma

APPLICANT: Shinkels, Richard A

APPLICANT: Herrman, John L

APPLICANT: Majumder, Kumud

APPLICANT: Mishra, Vishnu

APPLICANT: Mezes, Peter S

APPLICANT: MacDougall, John

TITLE OF INVENTION: No. US20020151151A1 Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 15966-697 CIP

CURRENT APPLICATION NUMBER: US/09/808,602

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 09/800,198

PRIOR FILING DATE: 2001-03-05

PRIOR APPLICATION NUMBER: 60/186,596

PRIOR FILING DATE: 2000-03-03

NUMBER OF SEQ ID NOS: 114

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 92

LENGTH: 2386

TYPE: DNA

ORGANISM: Homo sapiens

US-09-808-602-92

Alignment Scores:

Pred. No.: 0.337 Length: 2386

Score: 81.00 Matches: 15

Percent Similarity: 58.82% Conservative: 5

Best Local Similarity: 44.12% Mismatches: 14

Query Match: 34.03% Indels: 0

Gaps: 0

US-09-874-198-1 (1-41) x US-09-808-602-92 (1-2386)

QY 8 ProgluProvalPheGlyArgLeuAlaSerProgluTyrrAlaAsnAsp 27

DB 115 CCTCAGAGTATTGGGGAGGTGACTTCCCTCTGTCGCCAAGCCTTACCCCAACAC 174

QY 28 GlnGluArgArgrTrpThrLeuThrAlaProProgluTyrrArg 41

DB 175 TTTGAACAACACGTCGTATCACAGTCCCGACGGGATACAGG 216

RESULT 16

US-09-880-107-2256

Sequence 2256, Application US/09880107

Patent No. US20020142981A1

GENERAL INFORMATION:

APPLICANT: Horne, Darci T.

APPLICANT: Vockley, Joseph G.

APPLICANT: Scherf, Uwe

APPLICANT: Gene Logic, Inc.

TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

FILE REFERENCE: 44921-5028-WO

CURRENT APPLICATION NUMBER: US/09/880,107

PRIOR FILING DATE: 2001-06-14

PRIOR APPLICATION NUMBER: US 60/211,379

PRIOR FILING DATE: 2000-06-14

PRIOR APPLICATION NUMBER: US 60/237,054

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 3950

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2256

LENGTH: 2493

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Genbank Accession No. US20020142981A1 M14058

US-09-880-107-2256

Alignment Scores:

Pred. No.: 0.355 Length: 2493

Score: 81.00 Matches: 15



Percent Similarity: 58.828  
Best Local Similarity: 44.128  
Query Match: 34.038  
DB: 10  
Conservative: 5  
Matches: 14  
Indels: 0  
Gaps: 0

US-09-874-198-1 (1-41) x US-09-880-107-2256 (1-2493)

QY 8 ProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrrAlaAsnAsp 27

DB 127 CCTCAGAGATTATTGGGAGGAGTGACTCCCTCTGTGTTCCCAAGCCTTACCCCAACAC 186

QY 28 GlnGluArgArgTrpThrLeuThrAlaProProGlyTyrrArg 41

DB 187 TTGGAACAACACACTGTGATCAGTCCACGCGATACAG 228

RESULT 17  
US-10-044-090-613  
Sequence 613, Application US/10044090

Patent No. US20020137081A1

GENERAL INFORMATION:

APPLICANT: Olga Bandman

TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION

FILE REFERENCE: PA-0028 US

CURRENT APPLICATION NUMBER: US/10/044,090

CURRENT FILING DATE: 2002-01-09

NUMBER OF SEQ ID NOS: 850

SOFTWARE: PERL Program

SEQ ID NO 613

LENGTH: 2555

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

OTHER INFORMATION: Incyte ID No. US20020137081A1 1003386CB1

US-10-044-090-613

Alignment Scores:

Pred. No.: 0.365

Score: 81.00

Percent Similarity: 58.828

Best Local Similarity: 44.128

Query Match: 34.038

DB: 12

Length: 2555

Matches: 15

Conservative: 5

Mismatches: 14

Indels: 0

Gaps: 0

US-09-874-198-1 (1-41) x US-10-044-090-613 (1-2555)

8 ProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrrAlaAsnAsp 27

DB 183 CCTCAGAGATTATTGGGAGGAGTGACTCCCTCTGTGTTCCCAAGCCTTACCCCAAC 242

QY 28 GlnGluArgArgTrpThrLeuThrAlaProProGlyTyrrArg 41

DB 243 TTGGAACAACACACTGTGATCAGTCCACGCGATACAG 284

RESULT 18

US-09-960-352-3224

Sequence 3224, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengbing

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 3224

LENGTH: 354

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 14-LIB34-044-01-E1-D9  
US-09-960-352-3224

Alignment Scores:

Pred. No.: 0.049

Score: 80.00

Percent Similarity: 61.768

Best Local Similarity: 41.188

Query Match: 33.618

DB: 10

Length: 354

Matches: 14

Conservative: 7

Mismatches: 13

Indels: 0

Gaps: 0

US-09-874-198-1 (1-41) x US-09-960-352-3224 (1-354)

QY 8 ProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrrAlaAsnAsp 27

DB 133 CCTCAGAGCTCTATGGGAGGAGTGATGTCCCTGTGTTCCCAAGCCTTACCCCAACAT 192

QY 28 GlnGluArgArgTrpThrLeuThrAlaProProGlyTyrrArg 41

DB 193 TTGAGAGACACACTGTGATCAGTACGCGGATACAG 234

RESULT 19

US-09-960-352-8661

Sequence 8661, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengbing

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 8661

LENGTH: 392

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 37-LIB34-015-01-E1-B2

US-09-960-352-8661

Alignment Scores:

Pred. No.: 0.0552

Score: 80.00

Percent Similarity: 61.768

Best Local Similarity: 41.188

Query Match: 33.618

DB: 10

Length: 392

Matches: 14

Conservative: 7

Mismatches: 13

Indels: 0

Gaps: 0

US-09-874-198-1 (1-41) x US-09-960-352-8661 (1-392)

QY 8 ProGluProValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrrAlaAsnAsp 27

DB 89 CCTCAGAGCTCTATGGGAGGAGTGATGTCCCTGTGTTCCCAAGCCTTACCCCAACAT 148

QY 28 GlnGluArgArgTrpThrLeuThrAlaProProGlyTyrrArg 41

DB 149 TTGAGAGACACACTGTGATCAGTACGCGGATACAG 190

RESULT 20

US-09-960-352-10119

Sequence 10119, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengbing

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511.006/37-21(10298)C





US-08-866-650-4

Alignment Scores:

Pred. No.:	65.8	Length:	3919
Score:	68.00	Matches:	11
Percent Similarity:	65.52%	Conservative:	8
Best Local Similarity:	37.93%	Mismatches:	10
Query Match:	28.57%	Indels:	0
DB:	2	Gaps:	0

US-09-874-198-1 (1-41) x US-08-866-650-4 (1-3919)

QY 13 GLYArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTyr 32  
||| ::: ||||| ::||| :::: ||| :::: ||| |||  
Db 2994 GGCCCTATCCACCGACTCCCACTGCGCAGACAGTACCAAGCAGAAAGATGACTTGG 3053

QY 33 ThrLeuThralaProProGlyTyrArg 41  
::: ||| ||||| ::|||  
Db 3054 GAATATCAGCGCCACTCCGCGCACCGA 3080

US-09-021-287-4

Sequence 4, Application US/09021287

Patent No. 5981717

GENERAL INFORMATION:

APPLICANT: Greenspan, Daniel S

APPLICANT: Takahara, Kazuhiko

APPLICANT: Hoffman, Guy G

TITLE OF INVENTION: Mammalian Tolloid-Like Protein

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/021,287

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/866,650

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Berson, Bennett J

REGISTRATION NUMBER: 37094

REFERENCE/DOCKET NUMBER: 960296, 93839

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166

INFORMATION FOR SEQ. ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3919 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

MOLECULE TYPE: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: 648..3689

OTHER INFORMATION: /product= "human mTll protein"

US-09-021-287-4

Alignment Scores:

Pred. No.:	65.8	Length:	3919
------------	------	---------	------

Score:

Percent Similarity:	68.00	Matches:	11
Best Local Similarity:	37.93%	Conservative:	8
Query Match:	28.57%	Mismatches:	10
DB:	2	Indels:	0
		Gaps:	0

US-09-874-198-1 (1-41) x US-09-021-287-4 (1-3919)

QY 13 GLYArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTyr 32  
||| ::: ||||| ::||| :::: ||| :::: ||| |||  
Db 2994 GGCCCTATCCACCGACTCCCACTGCGCAGACAGTACCAAGCAGAAAGATGACTTGG 3053

QY 33 ThrLeuThralaProProGlyTyrArg 41  
::: ||| ||||| ::|||  
Db 3054 GAATATCAGCGCCACTCCGCGCACCGA 3080

RESULT 18

US-09-240-473-4

Sequence 4, Application US/09240473

Patent No. 6297011

GENERAL INFORMATION:

APPLICANT: Greenspan, Daniel S

APPLICANT: Takahara, Kazuhiko

APPLICANT: Hoffman, Guy G

TITLE OF INVENTION: Mammalian Tolloid-Like Protein

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/240,473

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Berson, Bennett J

REGISTRATION NUMBER: 37094

REFERENCE/DOCKET NUMBER: 960296, 93839

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166

INFORMATION FOR SEQ. ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3919 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

MOLECULE TYPE: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: 648..3689

OTHER INFORMATION: /product= "human mTll protein"

US-09-240-473-4

Alignment Scores:

Pred. No.:	65.8	Length:	3919
Score:	68.00	Matches:	11
Percent Similarity:	65.52%	Conservative:	8
Best Local Similarity:	37.93%	Mismatches:	10
Query Match:	28.57%	Indels:	0
DB:	4	Gaps:	0

US-09-874-198-1 (1-41) x US-09-240-473-4 (1-3919)

;; TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLOID-LIKE  
;; TITLE OF INVENTION: PROTEIN  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: RAINIER & PRESTIA  
;; STREET: P.O. BOX 980  
;; CITY: VALLEY FORGE  
;; STATE: PA  
;; COUNTRY: USA  
;; ZIP: 19482  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FASTSEQ for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/991,408  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 60/034,471  
;; FILING DATE: 02-JAN-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: PRESTIA, PAUL F.  
;; REGISTRATION NUMBER: 23,031  
;; REFERENCE/DOCKET NUMBER: ATG-50038  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 610-407-0700  
;; TELEFAX: 610-407-0701  
;; TELEX: 846169  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3690 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; US-08-991-408-3  
  
Alignment Scores:  
Pred. No.: 61.6 Length: 3690  
Score: 68.00 Matches: 11  
Percent Similarity: 65.52% Conservative: 8  
Best Local Similarity: 37.93% Mismatches: 10  
Query Match: 28.57% Indels: 0  
DB: 3 Gaps: 0  
  
US-09-874-198-1 (1-41) x US-08-991-408-3 (1-3690)  
13 GtAgtGleuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTrrp 32  
1143 GGCCTCATCCAGCTCCACACTGCGCCAGACAGACTACCCAGACGAGAGATGCACCTGG 1202  
DB 1143 GGCCTCATCCAGCTCCACACTGCGCCAGACAGACTACCCAGACGAGAGATGCACCTGG 1202  
  
QY 33 ThrLeuThrAlaProProGlyTyrArg 41  
Db 1203 GAATCAGCGCCACTCCTGCGCCACCGA 1229  
  
RESULT 15  
US-09-432-473-3  
;; Sequence 3, Application US/09432473  
;; Patent No. 6365715  
;; GENERAL INFORMATION:  
;; APPLICANT: ARLETH, ANTHONY J.  
;; APPLICANT: WILLETTTE, ROBERT N.  
;; APPLICANT: ELSHOURBAGY, MABIL A.  
;; APPLICANT: LI, XIAOTONG  
;; TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLOID-LIKE PROTEIN  
;; FILE REFERENCE: ATG-50038-D1  
;; CURRENT APPLICATION NUMBER: US/09/432,473  
;; CURRENT FILING DATE: 1999-11-01  
;; EARLIER APPLICATION NUMBER: 08/991,408  
;; EARLIER FILING DATE: 1997-12-16  
;; EARLIER APPLICATION NUMBER: 60/034,471

;; EARLIER FILING DATE: 1997-01-02  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FASTSEQ for Windows Version 3.0  
;; SEQ ID NO: 3  
;; LENGTH: 3690  
;; TYPE: DNA  
;; ORGANISM: HOMO SAPIENS  
;; US-09-432-473-3  
  
Alignment Scores:  
Pred. No.: 61.6 Length: 3690  
Score: 68.00 Matches: 11  
Percent Similarity: 65.52% Conservative: 8  
Best Local Similarity: 37.93% Mismatches: 10  
Query Match: 28.57% Indels: 0  
DB: 4 Gaps: 0  
  
US-09-874-198-1 (1-41) x US-09-432-473-3 (1-3690)  
QY 13 GtAgtGleuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTrrp 32  
Db 1143 GGCCTCATCCAGCTCCACACTGCGCCAGACAGACTACCCAGACGAGAGATGCACCTGG 1202  
DB 1203 GAATCAGCGCCACTCCTGCGCCACCGA 1229  
  
QY 33 ThrLeuThrAlaProProGlyTyrArg 41  
Db 1203 GAATCAGCGCCACTCCTGCGCCACCGA 1229  
  
RESULT 16  
US-08-866-650-4  
;; Sequence 4, Application US/08866650  
;; Patent No. 5939321  
;; GENERAL INFORMATION:  
;; APPLICANT: Greenspan, Daniel S  
;; APPLICANT: Takahara, Kazuhiko  
;; APPLICANT: Hoffman, Guy G  
;; TITLE OF INVENTION: Mammalian Tolloid-Like Protein  
;; NUMBER OF SEQUENCES: 13  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Quarles & Brady  
;; STREET: 1 South Pinckney Street  
;; CITY: Madison  
;; STATE: WI  
;; COUNTRY: US  
;; ZIP: 53703  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/866,650  
;; FILING DATE:  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Berson, Bennett J  
;; REGISTRATION NUMBER: 37094  
;; REFERENCE/DOCKET NUMBER: 960296, 93839  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 608-251-5000  
;; TELEFAX: 608-251-9166  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3919 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 648..3689  
;; OTHER INFORMATION: /product= "human mt11 protein"

Db	1798	GGCTCCATCATCACAGCGGGGCTGGGCCAAGAGATACCCCCCAACAAGACTGCTGG	1857
Qy	33	ThrlenthrlaProProglyTyrarg	41
Db	1858	CAGCTGGTGGCCCCCACACCAAGTACCGC	1884

RESULT 12  
US-08-377-292-1  
Sequence 1, Application US/08377292  
Patent No. 5693615  
GENERAL INFORMATION:  
APPLICANT: STONE, ROGER L.  
TITLE OF INVENTION: THERAPEUTIC FORMULAS FOR OSTEOINDUCTION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Procter & Gamble Company  
STREET: 11810 East Miami River Road  
CITY: Cincinnati  
STATE: Ohio  
COUNTRY: U.S.A.  
ZIP: 45239-8707  
COMPUTER READABLE FORM:

Alignment Scores:	
Pred. No.:	30.5
Score:	69.00
Percent Similarity:	58.6%
Best Local Similarity:	44.8%
Query Match:	28.9%
DB:	1
Length:	248
Matches:	13
Conservative:	4
Mismatches:	12
Indels:	0
Gaps:	0

US-09-874-198-1 (1-41) x US-08-377-292-1 (1-2487)

[illegible]

RESULT 13  
US-08-872-757-3  
; Sequence 3, Application US/08872757  
; Patent No. 6258584  
; GENERAL INFORMATION:

APPLICANT: Prockop, Darwin J.  
 APPLICANT: Hojima, Yoshio  
 APPLICANT: Li, Shi-Wu  
 APPLICANT: Stetson, Aleksander  
 TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND  
 TITLE OF INVENTION: PROCESSES, METHODS AND USES THEREOF  
 NUMBER OF SPOUNCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 City: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/872,757  
 FILING DATE: 10-JUN-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/609,187  
 FILING DATE: 01-MAR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Halluin, Albert P.  
 REGISTRATION NUMBER: 25,227  
 REFERENCE/DOCKET NUMBER: 8189-028-999

Pred. No.:	58.9	Length:	3546
Score:	68.00	Matches:	10
Percent Similarity:	65.5%	Conservative:	9
Best Local Similarity:	34.4%	Mismatches:	10
Query Match:	26.5%	Indels:	0
DB:	4	Gaps:	0

US-09-874-198-1 (1-41) X US-08-872-757-3 (1-3546)

Qy	3Th1euthTh1a1aProProgi1y1r1a1g 41
Db	2326 GGCATTCCTCCAGACCCCGGGCCACCG 2352
Qy	13 G1A1r1g1e1u1l1a1s1e1r1P1r1o1g1i1y1l1u1r1A1a1a1s1n1g1n1l1u1r1a1r1g1r1p 32
Db	2266 GGTACCATCTACACGCCCAACTGGCTGACAGTAATCCAGCAGAAAGAGATGACACTGG 2325

RESULT 14  
US-08-991-408-3  
Sequence 3, Application US/08991408  
Patient No. 608017  
GENERAL INFORMATION:  
APPLICANT: ARLETH, ANTHONY J.  
APPLICANT: WILLETTE, ROBERT N.  
APPLICANT: ELSHORBANGY, NABIL A.  
APPLICANT: LI, XIAORONG

Query Match:	29.41%	IndeIs:	6
DB:	2	Gaps:	2
US-09-874-198-1 (1-41) x US-08-400-159-7 (1-4464)			
QY	2	ProLeuGIyProLysTrPProGIuProValPheGIyArgLeuAlaSerProGIyPhePro	21
Db	1650	CCATGGGGCCACACACGCCGGAG-----GCTGCTGTGGCAGGATCCCA	1606
QY	22	GIyLysIyrrAlaAsnAspGIuArgArGIrPThrLeuThrAlaProProGIyIyrr	40
Db	1605	GGCCCC---GCGTGTGACCCGACGCCATGCATGACTCTGCAAGGCCCGCCAGGCGAC	1552

```

10
US-08-611-729A-7/c
Sequence 7, Application US/08611729A
Patent No. 6004924
GENERAL INFORMATION:
APPLICANT: Ish-Horowitz, David
APPLICANT: Henrique, Domingos M.P.
APPLICANT: Lewis, Julian H.
APPLICANT: Myat, Anna M.
APPLICANT: Fleming, Robert J.
APPLICANT: Aravanis-Tsakonas, Spyridon
APPLICANT: Mann, Robert S.
APPLICANT: Gray, Grace E.
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
NUMBER OF INVENTION: SRRATE GENE AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Penille & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611.729A
FILING DATE: 06-MAR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7526-037
ELECTRONIC COMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4483 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 332..4102
US-08-611-729A-7
Alignment Scores:
Pred. No.: 44.7 Length: 4483
Score: 70.00 Matches: 18
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 46.15% Mismatches: 12
Query Match: 29.41% Gaps: 6
DB: 3
2

```

US-09-874-198-1 (1-41) × US-08-611-729A-7 (1-4483)

Oy      2    ProLeuGlyProLysTrpPheGluProValPhcGlyArgLeuAlaSerProCylPhePro 21  
         |||     |||||       |||||       |||     |||||       |||  
Db    1650 CCATGGGGGCCACACAGCCGGAGA-----GCCTCTGTGCACAGCATCCA 1606

Oy      22 GlycIuTyrAlaAsnApSgInGUatGAryGrF<sup>rh</sup>leuThrAlaProCylYtyr 40  
         :       :       :       :       :       :       :       :  
db    1605 GGCCCC--GCGTCGACCCGACGCCATGCATCA<sup>mcyt</sup>TGCAGGCCCGCCGACAGGAC 1552

```

RESULT 11
US-08-872-757-1
Sequence 1, Application US/08872757
Patent No. 6258584
GENERAL INFORMATION:
APPLICANT: PROCKOP, Darwin J.
APPLICANT: Hojima, Yoshio
APPLICANT: Li, Shi-Wu
APPLICANT: Sieron, Aleksander
TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
TITLE OF INVENTION: PROCESSES, METHODS AND USES THEREOF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
City: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/872,757
FILING DATE: 10-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/609,187
FILING DATE: 01-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-028-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2457 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2190
US-08-872-757-1
Alignment Scores:
Pred. NO.: 30.1 Length: 2457
Score: 69.00 Matches: 13
Percent Similarity: 58.62% Conservative: 4
Best Local Similarity: 44.85% Mismatches: 12
Query Match: 28.99% Indels: 0
DB: 4 Gaps: 0
US-09-874-198-1 (1-41) x US-08-872-757-1 (1-2457)
OY 13 GlyArgLeuAlaSerProGlyIlePheProGlyIleTyrAlaAsnAspGingluarGArGTtp 32
||||| ::: ||||| :::::||| ||||| :::: |||

```

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,008
; FILING DATE: 23-APR-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/563,697
; FILING DATE: 28-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1506 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-839-008-6

Alignment Scores:
Pred. No.: 4.59 Length: 1506
Score: 74.00 Matches: 16
Percent Similarity: 48.89% Conservative: 6
Best Local Similarity: 35.56% Mismatches: 13
Query Match: 31.09% Indels: 10
DB: 2 Gaps: 1

US-09-874-198-1 (1-41) x US-08-839-008-6 (1-1506)

QY 5 ProlystPrProgluProvalPhe-----GlyArg 14
   ||| :: |||||
Db 122 CCCACTACGACGAGACTGTCTGTGCGGAGGAGCGTGCAGCGGAGTCAGGTTAC 181
   |||

QY 15 LeuAlaSerProlyPheProGlyGluTyrAlaAsnAspGlnGluArgGATGTTPhleu 34
   ::||| ||||| ||||| ||| ::|||
Db 182 GTGGCAAGTGGAGGTTCCCACTGTACCCCAACAAAGAGTCATCTGACAAATT 241
   ||| ||| |||
QY 35 ThrAlaProProGly 39
   ||| ||| |||
   242 ACGGTGCCCGAGGGG 256

      8
214-278-4/C
; Sequence 4, Application US/09214278
; Patent No. 6291210
; GENERAL INFORMATION:
; APPLICANT: Sakano, Sei-ji
; APPLICANT: Itoh, Akira
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8576
; CURRENT APPLICATION NUMBER: US/09/214,278
; CURRENT FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3955
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)..(3725)
; NAME/KEY: sig_peptide
; LOCATION: (12)..(89)
; NAME/KEY: mat_peptide
; LOCATION: (90)..(3725)
; US-09-214-278-4
```

```

Alignment Scores:
Pred. No.: 38.9 Length: 3955
Score: 70.00 Matches: 18
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 46.15% Mismatches: 12
Query Match: 29.41% Indels: 6
DB: 4 Gaps: 2

US-09-874-198-1 (1-41) x US-09-214-278-4 (1-3955)

QY 2 ProluGlyProlystPrProgluProvalPheGlyArgLeuAlaSerProGlyPhePro 21
   ||| ||||| ||||| ||| ||||| |||
Db 1840 CCATGGGGGCCACACAGCGCGAG-----GCTGCTGTGCCAGCATCCCA 1796
   |||

QY 22 GlyGluTyrAlaAsnAspGlnGluArgGATGTTPhleuThrAlaProProGlyTyr 40
   ||| ||| ||| ::||| ||||| ||||| ||||| ||||| |||||
Db 1795 GGCCCC---GCGTGTGACCGCGAGCCATCATCATCTGACAGCCCGCGAGGCGAC 1742
   |||

RESULT 9
US-08-400-159-7/C
; Sequence 7, Application US/08400159
; Patent No. 5869282
; GENERAL INFORMATION:
; APPLICANT: Ish-Horowitz, David
; APPLICANT: Henrique, Domingos M.P.
; APPLICANT: Lewis, Julian H.
; APPLICANT: Wyal, Anna M.
; APPLICANT: Fleming, Robert J.
; APPLICANT: Arlavanis-Tsakonas, Spyridon
; APPLICANT: Mann, Robert S.
; APPLICANT: Gray, Grace E.
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,159
; FILING DATE: 07-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mastrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4464 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-400-159-7

Alignment Scores:
Pred. No.: 44.5 Length: 4464
Score: 70.00 Matches: 18
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 46.15% Mismatches: 12
```



EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
ID NO: 37  
LENGTH: 985  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (633)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (642)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-227-357-37

Alignment Scores:  
Pred. No.: 1.68 Length: 985  
Score: 76.00 Matches: 15  
Percent Similarity: 62.07% Conservative: 3  
Best Local Similarity: 51.72% Mismatches: 11  
Query Match: 31.93% Indels: 0  
DB: 4 Gaps: 0

US-09-874-198-1 (1-41) x US-09-227-357-37 (1-985)

QY 12 PhgGLAArgLeuAlaSerProGlyPheProGlyGLuTYrAlaAsnAspGlnGluArgArg 31  
DB 690 TTGGCAAGCTGAGAGCCCTGGATGGCCAGATTAAGTACGACGACATGATGATGCNCC 631  
32 TrpThrLeuThrAlaProProGlyTyr 40  
630 GTTACTCTCAGACGCCGCCGACACACAY 604

RESULT 6  
US-08-839-008-4  
Sequence 4, Application US/08839008  
Patent No. 5916758  
GENERAL INFORMATION:  
APPLICANT: Hurle, Mark R  
APPLICANT: McDonnell, Peter C  
APPLICANT: McNulty, Dean E  
APPLICANT: Rosen, Craig A  
APPLICANT: Siemens, Ivo R  
APPLICANT: Young, Peter R  
APPLICANT: Yue, Tian-Li  
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/839,008  
FILING DATE: 23-APR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/563,697  
FILING DATE: 28-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: P50384  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1504 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-839-008-4

Alignment Scores:  
Pred. No.: 4.58 Length: 1504  
Score: 74.00 Matches: 16  
Percent Similarity: 48.89% Conservative: 6  
Best Local Similarity: 35.56% Mismatches: 13  
Query Match: 31.09% Indels: 10  
DB: 2 Gaps: 1

US-09-874-198-1 (1-41) x US-08-839-008-4 (1-1504)

QY 5 ProGlyTrpProGluProValPhe-----GlyArg 14  
DB 122 CCCAACTACAGAGACCTGTGTCCTGCGGAGGAGCTGACCGGGAGTCAGGTTAC 181  
QY 15 LeuAlaSerProGlyPheProGlyGLuTYrAlaAsnAspGlnGluArgTrpThrLeu 34  
DB 182 GTGGCAAGTAGGAGTTTCCCAACTCTACCCGCCAAGACAGATGATGTGACAAATT 241  
QY 35 ThrAlaProProGly 39  
DB 242 ACGGTGCCCGAGGGG 256

RESULT 7  
US-08-839-008-6  
Sequence 6, Application US/08839008  
Patent No. 5916758  
GENERAL INFORMATION:  
APPLICANT: Hurle, Mark R  
APPLICANT: McDonnell, Peter C  
APPLICANT: McNulty, Dean E  
APPLICANT: Rosen, Craig A  
APPLICANT: Siemens, Ivo R  
APPLICANT: Young, Peter R  
APPLICANT: Yue, Tian-Li  
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk



```

SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,350B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wolfe, Susan A
REGISTRATION NUMBER: 33,568
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4360 base pairs
type: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rattus rattus
FEATURE:
NAME/KEY: CDS
LOCATION: 94..3963
US-08-470-350B-1

Alignment Scores:
Pred. No.: 1.17 length: 4360
Score: 83.50 Matches: 16
Percent Similarity: 41.18 Conservative: 5
Best Local Similarity: 31.37% Mismatches: 17
Query Match: 35.08% Indels: 13
DB: 1 Gaps: 1

US-09-874-198-1 (1-41) x US-08-470-350B-1 (1-4360)

Oy 4 GlyProLysTrpProGluProValPhe----- 12
Db 2566 GGTCCACATTGGACACAGCTCTCCTTTGTAACACTATCTTGGAGAGCTTCTCGTACGGA 2625
Oy 13 -----GlyArgGluAlaSerProGlyPheProGlyGluTyrAlaAsnAspIleuLarg 30
Db 2626 CTCTCTGGGCAATTTCTGTACCCACTACTACCTGCGGAGGAGCTATCTTAATATGCGAGATGT 2685
Oy 31 ArgTrpThrLeuThrAlaProProGlyTyrArg 41
Db 2686 TTGTGGACATTGAGATCCCAACAACTACCGC 2718
US-09-839-008-8
Sequence 8, Application US/08839008
Patent No. 5916758
GENERAL INFORMATION:
APPLICANT: Hurtle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-Apr-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-839-008-8

Alignment Scores:
Pred. NO.: 0.691 Length: 1480
Score: 81.00 Matches: 17
Percent Similarity: 48.89% Conservative: 5
Best Local Similarity: 37.78% Mismatches: 13
Query Match: 34.03% Indels: 10
DB: 2 Gaps: 1

US-09-874-198-1 (1-41) x US-08-839-008-8 (1-1480)

QY 5 ProlystirProGluProValphe-----GlyArg 14
||| ::| ||||| ||| |||
Db 142 CCCACTCACCACACCCGCTGCTGCGGAGGAGTGAAGGGGAAATCAGTTAC 201
15 LeuAlasrProGlyPheProGlyGluTyrAlaAsnAspGingluArgTrrPThLeu 34
::| ||||| ||||| ||||| ||| :||| ||||| ::|
Db 202 GTGGCAATGTGGGGGTTCCCACTCCTACCCCCCTAATAAGAGTGCACTGTGACCATTA 261
QY 35 ThrAlaProProGly 39
||| ||| ||| |||
Db 262 ACGGTCCCCGAGGGC 276

RESULT 3
US-09-381-779-1
; Sequence 1, Application US/09381779
; Patent No. 6190868
; GENERAL INFORMATION:
; APPLICANT: Rothberg, Jonathan M.
; APPLICANT: Deem, Michael W.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING A NUCLEIC ACID SEQUENCE
; FILE REFERENCE: Cura-8 US 15966-508
; CURRENT APPLICATION NUMBER: US/09/381,779
; CURRENT FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 0.U.S.N. 60/054,887
; PRIOR FILING DATE: 1997-08-07
; PRIOR APPLICATION NUMBER: PCT/US98/16548
; PRIOR FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2492
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-381-779-1

Alignment Scores:

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: January 11, 2003, 11:09:54 ; Search time 74 Seconds  
(without alignments)  
169.916 Million cell updates/sec

Title: US-09-874-198-1  
Perfect score: 238  
Sequence: 1 TPLGPKWPEVPGRLASPGF.....GEVANDQERRWTLTAPPGR 41

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 60 summaries

Command line parameters:

-MODEL=frame+ p2n model -DRV=xld  
-Q/cgcn2\_1/USPTO.spool/US09874198/runat.10012003\_092044\_2595/app\_query.fasta.1.199  
-DB=Issued\_Patents.NA -QFMT=fastap -SUFIX=p2n.rni -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=60 -DOALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=20  
-MODE=LOCAL -OUFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09874198.cgcN1.1.31.0runat.10012003\_092044\_2595 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Issued\_Patents.NA.\*
- 2: /cgcn2\_6/ptodata/1/ina/5A.COMB.seq.\*
- 3: /cgcn2\_6/ptodata/1/ina/5B.COMB.seq.\*
- 4: /cgcn2\_6/ptodata/1/ina/5A.COMB.seq.\*
- 5: /cgcn2\_6/ptodata/1/ina/5B.COMB.seq.\*
- 6: /cgcn2\_6/ptodata/1/ina/Backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83.5	35.1	4360	1	US-08-470-350B-1
2	81	34.0	1480	1	US-08-839-008-8
3	81	34.0	2492	4	US-09-381-779-1
4	78	32.8	1537	2	US-08-839-008-1
5	76	31.9	985	4	US-09-227-357-37
6	74	31.1	1504	2	US-08-839-008-4
7	74	31.1	1506	2	US-08-839-008-6
8	70	29.4	3955	4	US-09-214-278-4
9	70	29.4	4464	2	US-08-400-159-7
10	70	29.4	4483	3	US-08-611-729A-7
11	69	29.0	2457	4	US-08-872-757-1
12	69	29.0	2487	1	US-08-377-292-1

13	68	28.6	3546	4	US-08-872-757-3	Sequence 3, Appl
14	68	28.6	3690	3	US-08-991-408-3	Sequence 3, Appl
15	68	28.6	3690	4	US-09-432-473-3	Sequence 3, Appl
16	68	28.6	3919	2	US-08-866-650-4	Sequence 4, Appl
17	68	28.6	3919	2	US-09-021-287-4	Sequence 4, Appl
18	68	28.6	3919	4	US-09-240-473-4	Sequence 4, Appl
19	68	28.6	5145	3	US-08-991-408-1	Sequence 1, Appl
20	68	28.6	5145	4	US-09-432-473-1	Sequence 1, Appl
21	67	28.2	4771	2	US-08-866-650-2	Sequence 2, Appl
22	67	28.2	4771	2	US-09-021-287-2	Sequence 2, Appl
23	67	28.2	4771	4	US-09-240-473-2	Sequence 2, Appl
24	65.5	27.5	1802	4	US-09-032-523-5	Sequence 5, Appl
25	64	26.9	3371	4	US-09-116-473-1	Sequence 1, Appl
26	63	26.5	1886	4	US-09-484-970B-127	Sequence 127, App
27	63	26.5	2730	3	US-08-936-135-17	Sequence 17, Appl
28	63	26.5	2766	3	US-08-936-135-3	Sequence 3, Appl
29	63	26.5	2772	3	US-08-936-135-1	Sequence 1, Appl
30	63	26.5	2781	3	US-08-936-135-19	Sequence 19, Appl
31	63	26.5	3471	4	US-09-116-473-3	Sequence 3, Appl
32	63	26.5	3652	3	US-08-936-135-5	Sequence 5, Appl
33	63	26.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl
34	62	26.1	2089	4	US-09-079-431B-5	Sequence 5, Appl
35	62	26.1	2254	4	US-09-079-431B-3	Sequence 3, Appl
36	62	26.1	2257	4	US-09-079-431B-1	Sequence 1, Appl
37	61.5	25.8	1636	4	US-09-039-198A-1	Sequence 1, Appl
38	61.5	25.8	1636	4	US-08-877-599-1	Sequence 1, Appl
39	61.5	25.8	1636	4	US-09-267-574-1	Sequence 1, Appl
40	61.5	25.8	1643	2	US-08-486-839-3	Sequence 3, Appl
41	61.5	25.8	1643	3	US-09-151-011-3	Sequence 3, Appl
42	61.5	25.8	1643	4	US-09-343-623-3	Sequence 3, Appl
43	61.5	25.8	1656	4	US-09-039-198A-3	Sequence 3, Appl
44	61.5	25.8	1656	4	US-08-877-599-3	Sequence 3, Appl
45	61.5	25.8	1656	4	US-09-267-574-3	Sequence 3, Appl
46	61.5	25.8	1713	2	US-08-486-839-5	Sequence 5, Appl
47	61.5	25.8	1713	3	US-09-151-011-5	Sequence 5, Appl
48	61.5	25.8	1713	4	US-09-343-623-5	Sequence 5, Appl
49	61	25.6	530	3	US-08-758-662-4	Sequence 4, Appl
50	61	25.6	838	1	US-07-590-894C-1	Sequence 1, Appl
51	61	25.6	3393	3	US-08-936-135-7	Sequence 7, Appl
52	61	25.6	4324	2	US-08-845-998-7	Sequence 7, Appl
53	61	25.6	4524	3	US-09-206-537-7	Sequence 7, Appl
54	61	25.6	4524	4	US-09-430-854-7	Sequence 7, Appl
55	61	25.6	4718	3	US-08-936-135-9	Sequence 9, Appl
56	61	25.6	4733	3	US-08-936-135-11	Sequence 11, Appl
57	61	25.6	4765	3	US-08-936-135-21	Sequence 21, Appl
58	61	25.6	4769	3	US-08-936-135-13	Sequence 13, Appl
59	61	25.6	4780	3	US-08-936-135-23	Sequence 23, Appl
60	61	25.6	4784	3	US-08-936-135-15	Sequence 15, Appl

#### ALIGNMENTS

RESULT 1  
US-08-470-350B-1  
; Sequence 1, Application US/08470350B  
; Patent No. 5684126  
GENERAL INFORMATION:  
APPLICANT: LI, Xiao  
TITLE OF INVENTION: Snyder, Solomon H  
TITLE OF INVENTION: Ebnerlin: A Secreted von Ebner's Gland  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus-p2n model

Run on: January 11, 2003, 11:09:24 ; Search time 2190 Seconds  
(without alignments)  
303.203 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPLGKWPPEVFGRLASPGF.....GEVANDQERMTLTAPPGYR 41

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Fgapop 10.0 , Fgapext 0.5	
Dgapop 6.0 , Dgapext 7.0	
Delop 6.0 , Delext 7.0	

16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 60 summaries

Command line parameters:  
-MODEL=frimex-p2n.model -DBV=xlp  
-O=/cgcn2.1/uspro.spool/US09874198/runat.10012003\_092044\_2571/app\_query.fasta.1.199  
-DB=EST -QFMT=fastap -SUFRTX=p2n.rst -MTNMATCH=0.1 -DOOPCL=0 -DOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -DIST=60  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=20 -MODE=LOCAL  
-OUTFMT=psco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09874198.cgcn.1.1.1716@runat.10012003\_092044\_2571 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-HARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1:	em_estba:*
2:	em_esthum:*
3:	em_estlu:*
4:	em_estlu:*
5:	em_estlov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hct:*
9:	gd_estl:*
10:	gd_estc:*
11:	gd_hct:*
12:	gd_estl:*
13:	gd_estl:*
14:	gd_estl:*
15:	em_estl:*
16:	em_estl:*
17:	gd_gss:*
18:	em_gss_hum:*
19:	em_gss_hum:*
20:	em_gss_hum:*
21:	em_gss_vit:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	238	100.0	796	14	B0650766	B0650766 AGENCOURT
2	238	100.0	911	14	B0654231	B0654231 AGENCOURT
3	238	100.0	913	14	B0648641	B0648641 AGENCOURT
4	238	100.0	921	14	B0648719	B0648719 AGENCOURT
5	238	100.0	939	14	B0644570	B0644570 AGENCOURT
6	238	100.0	975	14	B0646367	B0646367 AGENCOURT
7	238	100.0	1035	14	B0653181	B0653181 AGENCOURT
8	183	76.9	348	10	B0656455	B0656455 AGENCOURT
9	183	76.9	357	10	B0670802	B0670802 AGENCOURT
10	183	76.9	416	14	B0671372	B0671372 AGENCOURT
11	183	76.9	546	9	AA237253	AA237253 mx15807.r
12	183	76.9	552	9	AA238260	AA238260 mx15801.r
13	183	76.9	586	9	A1529228	A1529228 ut61809.y
14	183	76.9	600	12	BF532409	BF532409 602074450
15	183	76.9	699	10	B0653589	B0653589 B0653589
16	183	76.9	729	9	A1226038	A1226038 ut08608.y
17	183	76.9	752	12	BF233497	BF233497 602024289
18	183	76.9	843	13	B146955	B146955 602021589
19	183	76.9	884	12	BF236712	BF236712 602028006
20	177	74.4	523	9	A1327510	A1327510 mb27h04.y
21	177	74.4	523	9	A1327510	A1327510 mb27h04.y
22	173	72.7	450	9	AA208933	AA208933 mw71c03.r
23	170	71.4	678	12	BF233993	BF233993 602024479
24	153	64.3	487	9	A1048966	A1048966 uc77c07.y
25	139	52.5	620	14	B0678079	B0678079 APL1.9.CO
26	125	52.5	598	9	A1391397	A1391397 mw71c03.y
27	114	47.9	175	9	AA212846	AA212846 mw86a02.r
28	93	39.1	224	12	B0663462	B0663462 ut-M-BGO-
29	93	39.1	382	14	R11746	R11746 yf85e11.r1
30	93	39.1	387	14	T82204	T82204 y095f10.r1
31	93	39.1	389	10	B0672700	B0672700 B0672700
32	93	39.1	390	10	B0672630	B0672630 B0672630
33	93	39.1	409	14	R12521	R12521 yf55e09.r1
34	93	39.1	416	14	T97116	T97116 y641e01.r1
35	93	39.1	457	14	H15611	H15611 ym27e03.r1
36	93	39.1	550	14	H19143	H19143 ym50a01.r1
37	93	39.1	604	14	B0660916	B0660916 H4067C05-
38	93	39.1	619	9	A1099203	A1099203 ue40c11.y
39	93	39.1	621	9	AL134380	AL134380 DKE2D547C
40	93	39.1	621	10	B0621182	B0621182 B0621182
41	93	39.1	645	10	BE048313	BE048313 tz49h03.y
42	93	39.1	646	10	AW957587	AW957587 EST369657
43	93	39.1	648	10	B0659136	B0659136 B0659136
44	93	39.1	653	10	AW238809	AW238809 xb34e11.y
45	93	39.1	677	12	BF532377	BF532377 602074416
46	93	39.1	698	10	B0623493	B0623493 B0623493
47	93	39.1	711	10	AW954607	AW954607 EST366677
48	93	39.1	719	10	B0640327	B0640327 B0640327
49	93	39.1	731	14	B0647443	B0647443 ut-M-EGOP
50	93	39.1	734	14	B0771154	B0771154 ut-M-FIO-
51	93	39.1	794	13	B1330963	B1330963 602983262
52	93	39.1	894	13	B1913343	B1913343 603178822
53	92	38.7	496	13	B1318546	B1318546 fp71c02.y
54	91.5	38.4	530	13	B1775020	B1775020 467367.MA
55	91.5	38.4	742	9	A1460182	A1460182 cm72c03.x
56	90	37.8	503	14	B0739185	B0739185 pt41c12.y
57	89	37.4	551	12	B0234368	B0234368 daa40f02.
58	89	37.4	744	10	AW211385	AW211385 uc08d03.y
59	88.5	37.2	391	10	B0672811	B0672811 B0672811
60	87	36.6	360	14	D70256	D70256 CEUK101H3F

ALIGNMENTS

RESULT 1  
LOCUS B0650766 796 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT\_8347134 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6284012  
5', mRNA sequence.  
ACCESSION B0650766  
VERSION B0650766  
KEYWORDS EST.  
SOURCE B0650766.1 GI:21774938  
ORGANISM human.  
REFERENCE Homo sapiens  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL NIH-MGC http://mgi.nci.nih.gov/  
COMMENT 1 (bases 1 to 796)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2480 row: o column: 21  
High quality sequence stop: 505.  
Location/Qualifiers  
1. 796  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6284012"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 142 a 264 c 243 g 144 t 3 others  
ORIGIN  
Alignment Scores:  
Pied. No.: 6,39e-18 Length: 796  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0  
US-09-874-198-1 (1-41) x B0650766 (1-796)  
QY 1 ThProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
Db 77 ACCCCCTGGGCGCCGAAGTGGCTGAACTGTGTGGGGCGCTGGCATCCCCGGCTTT 136  
|||||  
QY 21 ProGlyGluTrpAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyr 40  
|||||  
Db 137 CCAGGGGAGATGCCAATGACAGGAGCGGCGCTGACCTGACTGACCCCGGCTTAC 196  
|||||  
QY 41 Arg 41  
|||  
Db 197 CGC 199

RESULT 2 B0654231 911 bp mRNA linear EST 15-JUL-2002  
LOCUS B0654231  
DEFINITION AGENCOURT\_8493289 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6299026

5', mRNA sequence.  
ACCESSION B0654231  
VERSION B0654231.1 GI:21778403  
KEYWORDS EST.  
SOURCE B0654231.1  
ORGANISM human.  
REFERENCE Homo sapiens  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL NIH-MGC http://mgi.nci.nih.gov/  
COMMENT 1 (bases 1 to 911)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2511 row: a column: 11  
High quality sequence start: 5  
High quality sequence stop: 461.  
Location/Qualifiers  
1. 911  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6299026"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 178 a 286 c 279 g 162 t 6 others  
ORIGIN  
Alignment Scores:  
Pied. No.: 7,35e-18 Length: 911  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0  
US-09-874-198-1 (1-41) x B0654231 (1-911)  
QY 1 ThProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
Db 77 ACCCCCTGGGCGCCGAAGTGGCTGAACTGTGTGGGGCGCTGGCATCCCCGGCTTT 136  
|||||  
QY 21 ProGlyGluTrpAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyr 40  
|||||  
Db 137 CCAGGGGAGATGCCAATGACAGGAGCGGCGCTGACCTGACTGACCCCGGCTTAC 196  
|||||  
QY 41 Arg 41  
|||  
Db 197 CGC 199

RESULT 3 B0648641 913 bp mRNA linear EST 15-JUL-2002  
LOCUS B0648641  
DEFINITION AGENCOURT\_8299192 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6271539  
5', mRNA sequence.  
ACCESSION B0648641  
VERSION B0648641.1 GI:21772813

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM2448 row: h column: 04  
High quality sequence stop: 549.  
Location/Qualifiers  
1. 913  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1ib="IMAGE:6271539"  
/clone\_1ib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pORF7; Site\_1: XhoI; Site\_2:  
EcoRI; CDNA made by oligo-dt priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-CDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."  
BASE COUNT 193 a 298 c 255 g 166 t 1 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.37e-18 Length: 913  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
-874-198-1 (1-41) x B0648641 (1-913)  
QY 1 ThnProLengIyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
DB 65 ACCCCCTTGGGCGCGAAGTGCTGAACCTGTGTGGGCGCGCTGGCATCCCCGGCTTT 124  
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr 40  
DB 125 CCAGGGGAGATGCATGACGAGGAGCGCGCTGCACCTGACACCCCGCGCTAC 184  
QY 41 Arg 41  
DB 185 CGC 187  
RESULT 4  
LOCUS B0648719 921 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT\_818306 NIH\_MGC\_100 Homo sapiens CDNA clone IMAGE:6283375  
5', mRNA sequence.  
ACCESSION B0648719  
VERSION B0648719.1 GI:21772891.  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM2479 row: e column: 08  
High quality sequence stop: 609.  
Location/Qualifiers  
1. 921  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1ib="IMAGE:6283375"  
/clone\_1ib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pORF7; Site\_1: XhoI; Site\_2:  
EcoRI; CDNA made by oligo-dt priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-CDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."  
BASE COUNT 172 a 300 c 285 g 164 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.44e-18 Length: 921  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0  
US-09-874-198-1 (1-41) x B0648719 (1-921)  
QY 1 ThnProLengIyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
DB 65 ACCCCCTTGGGCGCGAAGTGCTGAACCTGTGTGGGCGCGCTGGCATCCCCGGCTTT 124  
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr 40  
DB 125 CCAGGGGAGATGCATGACGAGGAGCGCGCTGCACCTGACACCCCGCGCTAC 184  
QY 41 Arg 41  
DB 185 CGC 187  
RESULT 5  
LOCUS B0644570 939 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT\_8490607 NIH\_MGC\_100 Homo sapiens CDNA clone IMAGE:6293770  
5', mRNA sequence.  
ACCESSION B0644570  
VERSION B0644570.1 GI:21768742  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE NIH-MGC <http://mhc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE

JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LNCM2497 row: f column: 11  
High quality sequence stop: 551.

## FEATURES

source

1. .939  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6293770"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pORF7; Site: 1: XhoI; Site: 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 189 a 308 c 273 g 166 t 3 others  
ORIGIN

## Alignment Scores:

Pred. No.: 7.59e-18 Length: 939  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-09-874-198-1 (1-41) x BQ644570 (1-939)

QY 1 ThProLeuGlyProLystrProGluProValPheGlyArgLeuAlaSerProGlyPhe 20

Db 65 ACCCCCTGGGGCCGAAGTGGCTGAACCTGTTCGGGGCCTGCACATCCCCGGCTTT 124

QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrThrLeuThrAlaProProGlyTyr 40

Db 125 CCAGGGGAGATGCCAATGACGAGCGGCGCTGCACCTGACCTGACCCCGGCTAC 184

41 Arg 41

Db 185 CGC 187

RESULT 6

LOCUS BQ646367 975 bp mRNA linear EST 15-JUL-2002

DEFINITION AGENCOURT\_8488545 NIH\_MGC\_100 Homo sapiens CDNA clone IMAGE:6296145

ACCESSION BQ646367

VERSION BQ646367.1 GI:21770539

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
NIH-MGC <http://mhc.nci.nih.gov/>.  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)

## FEATURES

source

CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LNCM2503 row: 1 column: 14  
High quality sequence stop: 504.

1. .975

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6296149"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pORF7; Site: 1: XhoI; Site: 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 198 a 320 c 286 g 171 t  
ORIGIN

## Alignment Scores:

Pred. No.: 7.9e-18 Length: 975  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-09-874-198-1 (1-41) x BQ646367 (1-975)

QY 1 ThProLeuGlyProLystrProGluProValPheGlyArgLeuAlaSerProGlyPhe 20

Db 65 ACCCCCTGGGGCCGAAGTGGCTGAACCTGTTCGGGGCCTGCACATCCCCGGCTTT 124

QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrThrLeuThrAlaProProGlyTyr 40

Db 125 CCAGGGGAGATGCCAATGACGAGCGGCGCTGCACCTGACCTGACCCCGGCTAC 184

41 Arg 41

Db 185 CGC 187

RESULT 7

LOCUS BQ653181 1035 bp mRNA linear EST 15-JUL-2002

DEFINITION AGENCOURT\_8303990 NIH\_MGC\_100 Homo sapiens CDNA clone IMAGE:6269223

ACCESSION BQ653181

VERSION BQ653181.1 GI:21777353

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
NIH-MGC <http://mhc.nci.nih.gov/>.  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be



found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: L1C2442 row: 9 column: 16  
High quality sequence stop: 510.

## FEATURES

source

Location/Qualifiers  
1..1035  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6269223"  
/clone\_lib="NIH\_MGC.100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."  
BASE COUNT 224 a 341 c 298 g 172 t  
ORIGIN

## Alignment Scores:

Pred. No.: 8,41e-18 Length: 1035  
Score: 238.00 Matches: 41  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-09-874-198-1 (1-41) x BQ653181 (1-1035)

QY 1 ThProleuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
65 ACCCCCTTGGGCGGAGTGGCTGAACTGTTCGGGGCCCTGGCATCCCCGGCTT 124  
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTyrPheThrAlaProProGlyTyr 40  
|||||  
Db 125 CCAGGGGAGATGCAATGACGAGGAGCGCCCTGCACCTGCATGACCCCGGCTAC 184  
QY 41 Arg 41  
|||  
Db 185 CGC 187

LOC 8  
425

AA896425 238 bp mRNA linear EST 06-APR-1998  
v33e05.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone  
IMAGE:1297088 5' similar to TR:000187 000187 MASP-2 PROTEIN. ;  
mRNA sequence.

ACCESSION AA896425  
VERSION AA896425.1 GI:3032818  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 238)  
AUTHORS Marmè, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Waters, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Maria W/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:678136  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 197.

## FEATURES

source

Location/Qualifiers  
1..238  
/organism="Mus musculus"  
/strain="C57BL/6 x CBA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1297088"  
/clone\_lib="Stratagene mouse lung 937302"  
/sex="female"  
/tissue\_type="lung"  
/dev\_stage="6-8 month old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. 6-8 month old female lung and 1.5 year old male lung were source of mRNA. Average insert size: 1.5 kb. Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."  
BASE COUNT 44 a 69 c 67 g 58 t  
ORIGIN

## Alignment Scores:

Pred. No.: 4.26e-12 Length: 238  
Score: 183.00 Matches: 32  
Percent Similarity: 85.37% Conservative: 3  
Best Local Similarity: 78.05% Mismatches: 6  
Query Match: 76.89% Indels: 0  
DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x AA896425 (1-238)

QY 1 ThProleuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20  
|||||  
Db 116 ACACCTTGGGCTTCAAGATGCTGAACTGTATTCGGGGCCCTGGTGTACCTGGCTTC 175  
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTyrPheThrAlaProProGlyTyr 40  
|||||  
Db 176 CCAGAGAAATGATGCTGACATCAAGATGATCTGCAGACTGACACCCCGGATAC 235  
QY 41 Arg 41  
|||  
Db 236 CGC 238

RESULT 9 348 bp mRNA linear EST 27-NOV-2001  
BB870802  
LOCUS BB870802  
DEFINITION BB870802 RIKEN full-length enriched, pooled tissues, intestinal mucosa, etc. Mus musculus cDNA clone G630025E18 5', mRNA sequence.  
ACCESSION BB870802  
VERSION BB870802.1 GI:17117012  
KEYWORDS EST.

SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 348)  
AUTHORS Akimura, T., Arakawa, T., Carinci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Soabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Watanishi, A., Yasunishi, A., Yamatsuta, M. and Hayashizaki, Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)  
JOURNAL Unpublished (2001)



```
/organism="Mus musculus"
/db_xref="taxon:10090"
```









ACCESSION mRNA sequence.  
 VERSION BF236712  
 KEYWORDS BF236712.1 GI:1150598  
 EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 884)  
 NIH-MGC <http://mgc.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM9447 row: k column: 01  
 High quality sequence stop: 615.  
 Location/Qualifiers

## FEATURES

source  
 1..884  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4163448"  
 /clone\_lib="NCI-CGAP\_L19"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.9 kb. Constructed by Life  
 Technologies. Note: this is a NCI-CGAP Library."  
 BASE COUNT 191 a 256 c 252 g 185 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	1.68e-11	Length:	884
Score:	183.00	Matches:	32
Percent Similarity:	85.378	Conservative:	3
Best Local Similarity:	78.058	Mismatches:	6
Query Match:	76.89%	Indels:	0
DB:	12	Gaps:	0

US-09-874-198-1 (1-41) x BF236712 (1-884)

```

1  TThProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
32  ACACCTCTGTGGTTCACAAAGCTGCTGAACCTGATTCGGGGCGCTGGTGCCTGCGCTT 91
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  21  ProGlyGluTrpAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyr 40
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    92  CCAGAGAAGATGCTGACCATCAAGATCGATCCTGGACACTGACGACCCCTGGCTAC 151
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY    41  Arg 41
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    152  CGC 154
```

Search completed: January 11, 2003, 12:46:28  
 Job time : 2197 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: January 11, 2003, 11:06:19 ; Search time 351 Seconds

(without alignments)  
263.054 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238  
Sequence: 1 TPLGPKWPEYFGRLASPGF.....GEYANDQERWTLTAPGVR 41

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Matched: 2165239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 60 summaries

Command line parameters:

-MODEL=frame\_plus.p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09874199/runat.10012003.092043.2552/app\_query.fasta.1.199  
-DB=N.Geneseq.101002 -OPMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=60 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=20  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09874198.ecgn.1.1.79.6runat.10012003.092043.2552 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	238	100.0	2472	24	AAD24224	Human MASP-2 CDNA
2	238	100.0	3106	24	ASB85321	Human encoding novel
3	93.5	39.3	810	22	ABA06477	Human CDNA SEQ ID
4	93.5	39.3	810	22	AA531599	CDNA encoding novel
5	93.5	39.3	810	22	ABK43587	CDNA encoding novel
6	93	39.1	363	21	AA001735	Human secreted pro
7	93	39.1	397	21	AAC00227	Human secreted pro
8	93	39.1	363	22	AAH43512	CDNA encoding prot
9	93	39.1	385	22	AAE83985	Human serine prote
10	91.5	38.4	1988	22	AAE74442	Human PRO11 nucleo
11	91.5	38.4	2127	22	AAE74440	Human PRO9 nucleo
12	91.5	38.4	2127	22	AAE74441	Human PRO10 nucleo
13	91.5	38.4	2143	22	AAE74443	Human PRO12 nucleo
14	91.5	38.4	3663	22	AAE74438	Human PRO7 nucleot
15	91.5	38.4	3879	22	AAE74439	Human PRO8 nucleot
16	91.5	38.4	4034	24	ABK28635	Human CDNA encodin
17	91.5	38.4	4198	24	AAD25344	Human SE26 DNA. H
18	84	35.3	9455	23	ABL10283	Drosophila melanog
19	84	35.3	13036	23	ABL10282	Drosophila melanog
20	83.5	35.1	4344	24	ABK63805	Rat sequence diffe
21	83.5	35.1	4360	18	AAT44068	Rat von Ebner's gl
22	81	34.0	1148	21	AAE18149	Lung cancer associ
23	81	34.0	1480	24	ABN95698	Gene #2196 used to
24	81	34.0	1480	24	ABK64507	Human benign prost
25	81	34.0	1480	24	ABK35519	Human endometrial
26	81	34.0	2492	20	AAK24297	Human complement c
27	81	34.0	2493	22	ABN83114	Complement compone
28	81	34.0	2493	24	ABN95758	Gene #2256 used to
29	79	33.2	382	21	AAA45208	Human secreted exp
30	79	33.2	2070	22	AA500132	Human CDNA encodin
31	79	33.2	2073	22	AA500128	Human CDNA encodin
32	79	33.2	2073	22	AA500136	Human CDNA encodin
33	79	33.2	2154	22	AA500116	Human CDNA encodin
34	79	33.2	2157	22	AA500112	Human CDNA encodin
35	79	33.2	2157	22	AA500120	Human CDNA encodin
36	79	33.2	3252	22	AA500133	Human CDNA encodin
37	79	33.2	3255	22	AA500129	Human CDNA encodin
38	79	33.2	3255	22	AA500137	Human CDNA encodin
39	79	33.2	3336	22	AA500117	Human CDNA encodin
40	79	33.2	3339	22	AA500113	Human CDNA encodin
41	79	33.2	3339	22	AA500121	Human CDNA encodin
42	79	33.2	3579	22	AA500138	Human CDNA encodin
43	79	33.2	3582	22	AA500127	Human CDNA encodin
44	79	33.2	3582	22	AA500135	Human CDNA encodin
45	79	33.2	3633	22	AA500115	Human CDNA encodin
46	79	33.2	3666	22	AA500111	Human CDNA encodin
47	79	33.2	3666	22	AA500119	Human CDNA encodin
48	79	33.2	3666	22	AA500130	Human CDNA encodin
49	79	33.2	3669	22	AA500126	Human CDNA encodin
50	79	33.2	3669	22	AA500134	Human CDNA encodin
51	79	33.2	3750	22	AA500114	Human CDNA encodin
52	79	33.2	3753	22	AA500110	Human CDNA encodin
53	79	33.2	3753	22	AA500118	Human CDNA encodin
54	79	33.2	4446	19	AA500131	Human CDNA encodin
55	78	32.8	1482	22	AAV22668	Nucleic acid encod
56	78	32.8	1537	18	AAT72652	Human smooth muscl
57	78	32.8	1537	20	AAK90357	Human smooth muscl
58	78	32.8	1542	24	ABK35135	Human CDNA encodin
59	78	32.8	1580	21	AAC78046	Human cancer assoc
60	78	32.8	1768	22	AAK94126	Human full-length

## ALIGNMENTS

RESULT 1  
AAD24224  
ID AAD24224 standard; cDNA: 2472 BP.



CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 3106 BP; 848 A; 846 C; 798 G; 614 T; 0 other;

## Alignment Scores:

Pred. No.:	1,36e-18	Length:	3106
Score:	238.00	Matches:	41
Global Similarity:	100.00%	Conservative:	0
Local Similarity:	100.00%	Mismatches:	0
Match:	100.00%	Indels:	0
	23	Gaps:	0

US-09-874-198-1 (1-41) x AAS85321 (1-3106)

OY 1 TheProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20

DB 2255 ACCCCCTTGGCCCGAAGTGGCTGAACCTGTTCGGCGCCTGGCATCCCGGCTTT 2314

OY 21 ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrThrLeuThrAlaProGlyTyr 40

DB 2315 CCAGGGAGATGCCAATGACACGAGCGGCGCTGACCTGACCTGACCCCGGCTTAC 2374

OY 41 Arg 41

DB 2375 CGC 2377

DB 2375 CGC 2377

RESULT 3  
ABA06477/c  
ID ABA06477 standard; cDNA: 810 BP.

XX ABA06477;

DT 10-JAN-2002 (first entry)

XX Human cDNA SEQ ID NO: 143.

DE Human; gene therapy; neural disorder; immune system disorder;  
muscular disorder; reproductive disorder; gastrointestinal disorder;  
pulmonary disorder; cardiovascular disorder; renal disorder;  
proliferative disorder; inflammation; ss.

XX Homo sapiens.

OS WO200154474-A2.

XX 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US01349.

XX 31-JAN-2000; 2000US-179065P.

XX 04-FEB-2000; 2000US-180628P.

XX 24-FEB-2000; 2000US-184664P.

XX 02-MAR-2000; 2000US-186350P.

XX 16-MAR-2000; 2000US-189874P.

XX 17-MAR-2000; 2000US-190076P.

XX 18-APR-2000; 2000US-198123P.

XX 19-MAY-2000; 2000US-205515P.

XX 07-JUN-2000; 2000US-209467P.

XX 28-JUN-2000; 2000US-214886P.

XX 30-JUN-2000; 2000US-215135P.

XX 07-JUL-2000; 2000US-216647P.

XX 07-JUL-2000; 2000US-216880P.

PR 11-JUL-2000; 2000US-217487P.  
PR 11-JUL-2000; 2000US-217496P.  
PR 14-JUL-2000; 2000US-218290P.  
PR 26-JUL-2000; 2000US-220963P.  
PR 26-JUL-2000; 2000US-220964P.  
PR 14-AUG-2000; 2000US-224518P.  
PR 14-AUG-2000; 2000US-224519P.  
PR 14-AUG-2000; 2000US-225213P.  
PR 14-AUG-2000; 2000US-225214P.  
PR 14-AUG-2000; 2000US-225265P.  
PR 14-AUG-2000; 2000US-225267P.  
PR 14-AUG-2000; 2000US-225268P.  
PR 14-AUG-2000; 2000US-225270P.  
PR 14-AUG-2000; 2000US-225447P.  
PR 14-AUG-2000; 2000US-225757P.  
PR 14-AUG-2000; 2000US-225758P.  
PR 14-AUG-2000; 2000US-225759P.  
PR 18-AUG-2000; 2000US-226279P.  
PR 22-AUG-2000; 2000US-226681P.  
PR 22-AUG-2000; 2000US-226686P.  
PR 22-AUG-2000; 2000US-227182P.  
PR 23-AUG-2000; 2000US-227009P.  
PR 30-AUG-2000; 2000US-228924P.  
PR 01-SEP-2000; 2000US-229287P.  
PR 01-SEP-2000; 2000US-229343P.  
PR 01-SEP-2000; 2000US-229344P.  
PR 05-SEP-2000; 2000US-229309P.  
PR 05-SEP-2000; 2000US-229513P.  
PR 06-SEP-2000; 2000US-230437P.  
PR 06-SEP-2000; 2000US-230438P.  
PR 08-SEP-2000; 2000US-231242P.  
PR 08-SEP-2000; 2000US-231243P.  
PR 08-SEP-2000; 2000US-231244P.  
PR 08-SEP-2000; 2000US-231413P.  
PR 08-SEP-2000; 2000US-231414P.  
PR 08-SEP-2000; 2000US-232080P.  
PR 08-SEP-2000; 2000US-232081P.  
PR 12-SEP-2000; 2000US-231968P.  
PR 14-SEP-2000; 2000US-232387P.  
PR 14-SEP-2000; 2000US-232388P.  
PR 14-SEP-2000; 2000US-232399P.  
PR 14-SEP-2000; 2000US-232400P.  
PR 14-SEP-2000; 2000US-232401P.  
PR 14-SEP-2000; 2000US-233063P.  
PR 14-SEP-2000; 2000US-233064P.  
PR 14-SEP-2000; 2000US-233065P.  
PR 21-SEP-2000; 2000US-234223P.  
PR 21-SEP-2000; 2000US-234274P.  
PR 25-SEP-2000; 2000US-234997P.  
PR 25-SEP-2000; 2000US-234998P.  
PR 26-SEP-2000; 2000US-235484P.  
PR 27-SEP-2000; 2000US-235834P.  
PR 27-SEP-2000; 2000US-235836P.  
PR 29-SEP-2000; 2000US-236337P.  
PR 29-SEP-2000; 2000US-236367P.  
PR 29-SEP-2000; 2000US-236368P.  
PR 29-SEP-2000; 2000US-236369P.  
PR 29-SEP-2000; 2000US-236370P.  
PR 02-OCT-2000; 2000US-237037P.  
PR 02-OCT-2000; 2000US-237038P.  
PR 02-OCT-2000; 2000US-237039P.  
PR 02-OCT-2000; 2000US-237040P.  
PR 13-OCT-2000; 2000US-239935P.  
PR 13-OCT-2000; 2000US-239937P.  
PR 20-OCT-2000; 2000US-240960P.  
PR 20-OCT-2000; 2000US-241221P.  
PR 20-OCT-2000; 2000US-241785P.  
PR 20-OCT-2000; 2000US-241786P.  
PR 20-OCT-2000; 2000US-241787P.  
PR 20-OCT-2000; 2000US-241808P.  
PR 20-OCT-2000; 2000US-241809P.

20-OCT-2000; 2000US-241826P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 08-NOV-2000; 2000US-246474P.  
PR 08-NOV-2000; 2000US-246475P.  
PR 08-NOV-2000; 2000US-246476P.  
PR 08-NOV-2000; 2000US-246477P.  
PR 08-NOV-2000; 2000US-246478P.  
PR 08-NOV-2000; 2000US-246523P.  
PR 08-NOV-2000; 2000US-246524P.  
PR 08-NOV-2000; 2000US-246525P.  
PR 08-NOV-2000; 2000US-246526P.  
PR 08-NOV-2000; 2000US-246527P.  
PR 08-NOV-2000; 2000US-246528P.  
PR 08-NOV-2000; 2000US-246532P.  
PR 08-NOV-2000; 2000US-246609P.  
PR 08-NOV-2000; 2000US-246610P.  
PR 08-NOV-2000; 2000US-246611P.  
PR 08-NOV-2000; 2000US-246613P.  
PR 17-NOV-2000; 2000US-249207P.  
PR 17-NOV-2000; 2000US-249208P.  
PR 17-NOV-2000; 2000US-249209P.  
PR 17-NOV-2000; 2000US-249210P.  
PR 17-NOV-2000; 2000US-249211P.  
PR 17-NOV-2000; 2000US-249212P.  
PR 17-NOV-2000; 2000US-249213P.  
PR 17-NOV-2000; 2000US-249214P.  
PR 17-NOV-2000; 2000US-249215P.  
PR 17-NOV-2000; 2000US-249216P.  
PR 17-NOV-2000; 2000US-249217P.  
PR 17-NOV-2000; 2000US-249218P.  
PR 17-NOV-2000; 2000US-249244P.  
PR 17-NOV-2000; 2000US-249245P.  
PR 17-NOV-2000; 2000US-249264P.  
PR 17-NOV-2000; 2000US-249265P.  
PR 17-NOV-2000; 2000US-249287P.  
PR 17-NOV-2000; 2000US-249299P.  
PR 17-NOV-2000; 2000US-249300P.  
PR 01-DEC-2000; 2000US-250160P.  
PR 01-DEC-2000; 2000US-250319P.  
PR 05-DEC-2000; 2000US-251030P.  
PR 05-DEC-2000; 2000US-251988P.  
PR 05-DEC-2000; 2000US-256719P.  
PR 06-DEC-2000; 2000US-251479P.  
PR 08-DEC-2000; 2000US-251856P.  
PR 08-DEC-2000; 2000US-251868P.  
PR 08-DEC-2000; 2000US-251989P.  
PR 08-DEC-2000; 2000US-251990P.  
PR 11-DEC-2000; 2000US-254097P.  
PR 05-JAN-2001; 2001US-259678P.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-476161/51.  
XX P-PSDB: ABB10255.  
XX  
XX Isolated nucleic acid molecule encoding an inflammation-associated  
PT polypeptide is used in preventing, treating or ameliorating a medical  
PT condition  
XX  
XX Claim 1, SEQ ID NO: 143; 859pp + Sequence Listing; English.  
XX  
XX The present invention provides human cDNAs, proteins and related genomic  
CC DNAs. These can be used in the treatment of neural, immune system,  
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,  
CC renal and proliferative disorders and inflammation. The present sequence  
CC is a cDNA of the invention.  
XX  
SQ Sequence 810 BP; 181 A; 189 C; 249 G; 187 T; 4 other;

Pred. No.: 0.0402  
Score: 93.50  
Percent Similarity: 46.154  
Best Local Similarity: 40.388  
Query Match: 39.298  
DB: 22  
US-09-874-198-1 (1-41) x ABA06477 (1-810)  
QY 5 ProllystP-----ProgluProval----- 11  
DB 606 CCCWCTGGGATTCAAGGACCCGTMGCATGCTTCGCGCGAGTGATCCCAAT 547  
QY 12 -----PheglYargLeuAlaSerProglYpHeProglYglutYrAlaAsnAspGln 29  
DB 546 GCCACCAACGCGCGCATGCTCTCCAGGCTTCGCGGACACACACACACCTAC 487  
QY 30 ArgArgTrpThrLeuThrAlaProProglYTrArg 41  
DB 486 TGTCACTGGCTCCTGAGGCTCTGAGGCGCACCG 451  
RESULT 4  
AAS31599/C  
ID AAS31599 standard; cDNA; 810 BP.  
XX  
AC AAS31599;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE cDNA encoding novel human calcium-binding protein #23.  
XX  
XX Human: calcium-binding protein; calcium flux; neurological disease;  
KW immune dysfunction; digestive disorder; neoplastic disease;  
KW blood disorder; infectious disease; gene therapy; immunosuppressive;  
KW antidiabetic; cytostatic; vasotropic; antibacterial; nootropic;  
KW virulence; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155304-A2.  
XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01302.  
PE  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.

```

PR 14-AUG-2000: 2000US-0225758.
PR 14-AUG-2000: 2000US-0225759.
PR 18-AUG-2000: 2000US-0226279.
PR 22-AUG-2000: 2000US-0226681.
PR 22-AUG-2000: 2000US-0226686.
PR 22-AUG-2000: 2000US-0227182.
PR 23-AUG-2000: 2000US-0227009.
PR 30-AUG-2000: 2000US-0228924.
PR 01-SEP-2000: 2000US-0229287.
PR 01-SEP-2000: 2000US-0229343.
PR 01-SEP-2000: 2000US-0229344.
PR 01-SEP-2000: 2000US-0229345.
PR 05-SEP-2000: 2000US-0229509.
PR 05-SEP-2000: 2000US-0229513.
PR 06-SEP-2000: 2000US-0230437.
PR 06-SEP-2000: 2000US-0230438.
PR 08-SEP-2000: 2000US-0231242.
PR 08-SEP-2000: 2000US-0231243.
PR 08-SEP-2000: 2000US-0231244.
PR 08-SEP-2000: 2000US-0231413.
PR 08-SEP-2000: 2000US-0231414.
PR 08-SEP-2000: 2000US-0232080.
PR 08-SEP-2000: 2000US-0232081.
PR 12-SEP-2000: 2000US-0231968.
PR 14-SEP-2000: 2000US-0232397.
PR 14-SEP-2000: 2000US-0232398.
PR 14-SEP-2000: 2000US-0232399.
PR 14-SEP-2000: 2000US-0232400.
PR 14-SEP-2000: 2000US-0232401.
PR 14-SEP-2000: 2000US-0233063.
PR 14-SEP-2000: 2000US-0233064.
PR 14-SEP-2000: 2000US-0233065.
PR 21-SEP-2000: 2000US-0234223.
PR 21-SEP-2000: 2000US-0234274.
PR 25-SEP-2000: 2000US-0234997.
PR 25-SEP-2000: 2000US-0234998.
PR 26-SEP-2000: 2000US-0235484.
PR 27-SEP-2000: 2000US-0235834.
PR 27-SEP-2000: 2000US-0235836.
PR 29-SEP-2000: 2000US-0236327.
PR 29-SEP-2000: 2000US-0236367.
PR 29-SEP-2000: 2000US-0236368.
PR 29-SEP-2000: 2000US-0236369.
PR 29-SEP-2000: 2000US-0236370.
PR 02-OCT-2000: 2000US-0236802.
PR 02-OCT-2000: 2000US-0237037.
PR 02-OCT-2000: 2000US-0237038.
PR 02-OCT-2000: 2000US-0237039.
PR 02-OCT-2000: 2000US-0237040.
PR 13-OCT-2000: 2000US-0239935.
PR 13-OCT-2000: 2000US-0239937.
PR 20-OCT-2000: 2000US-0240960.
PR 20-OCT-2000: 2000US-0241221.
PR 20-OCT-2000: 2000US-0241785.
PR 20-OCT-2000: 2000US-0241786.
PR 20-OCT-2000: 2000US-0241787.
PR 20-OCT-2000: 2000US-0241808.
PR 20-OCT-2000: 2000US-0241809.
PR 20-OCT-2000: 2000US-0241826.
PR 01-NOV-2000: 2000US-0244617.
PR 08-NOV-2000: 2000US-0246474.
PR 08-NOV-2000: 2000US-0246475.
PR 08-NOV-2000: 2000US-0246476.
PR 08-NOV-2000: 2000US-0246477.
PR 08-NOV-2000: 2000US-0246478.
PR 08-NOV-2000: 2000US-0246523.
PR 08-NOV-2000: 2000US-0246524.
PR 08-NOV-2000: 2000US-0246525.
PR 08-NOV-2000: 2000US-0246526.
PR 08-NOV-2000: 2000US-0246527.
PR 08-NOV-2000: 2000US-0246528.
PR 08-NOV-2000: 2000US-0246532.
PR 08-NOV-2000: 2000US-0246609.

PR 08-NOV-2000: 2000US-0246610.
PR 08-NOV-2000: 2000US-0246611.
PR 08-NOV-2000: 2000US-0246613.
PR 17-NOV-2000: 2000US-0249207.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249209.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249212.
PR 17-NOV-2000: 2000US-0249213.
PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249244.
PR 17-NOV-2000: 2000US-0249245.
PR 17-NOV-2000: 2000US-0249264.
PR 17-NOV-2000: 2000US-0249265.
PR 17-NOV-2000: 2000US-0249297.
PR 17-NOV-2000: 2000US-0249299.
PR 17-NOV-2000: 2000US-0249300.
PR 01-DEC-2000: 2000US-0250160.
PR 01-DEC-2000: 2000US-0250391.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 05-DEC-2000: 2000US-0256719.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251899.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
PR 03-JAN-2001: 2001US-0259578.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM.
XX WPI; 2001-465568/50.
DR P-PSDB; AAU19914.
XX
XX
PT Isolated nucleic acid molecule encoding a calcium-binding protein 1s
PT used in preventing, treating or ameliorating a medical condition
XX
XX
PS Claim 4; SEQ ID NO 33; 542pp; English.
XX
XX
CC The present invention relates to the isolation of novel human
CC calcium-binding proteins (AAU19892-AAU19969) and cDNA and genomic
CC sequences encoding for these proteins. The sequences of the invention
CC are useful in the diagnosis, prevention and/or prognosis of diseases
CC associated with aberrant calcium flux. Such disorders include
CC neurological diseases (e.g. amyotrophic lateral sclerosis, ALS),
CC immune dysfunction (e.g. severe combined immunodeficiency, SCID),
CC digestive disorders (e.g. irritable bowel syndrome, IBS), neoplastic
CC disease (e.g. cancer), blood disorders (e.g. haemophilia), and/or
CC infectious disease (e.g. acquired immunodeficiency syndrome, AIDS). The
CC novel calcium-binding proteins are also useful as screening tools to
CC identify antagonists and/or agonists that may enhance or inhibit
CC activities mediated by calcium-binding proteins. The polynucleotides of
CC the invention are also useful in gene therapy. AAS3157-AAS31654
CC represent cDNA sequences encoding for the novel human calcium-binding
CC proteins.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SO Sequence 810 BP; 181 A; 189 C; 249 G; 187 T; 4 other:

Alignment Scores:
Pred. No.: 0.0402 Length: 810
Score: 93.50 Matches: 21

```

Percent Similarity: 46.15%  
 Best Local Similarity: 40.38%  
 Query Match: 39.29%  
 DB: 22  
 Gaps: 2

US-09-874-198-1 (1-41) x AAS31599 (1-810)

QY 5 ProlytP-----ProgiuProval----- 11  
 DB 606 CCCWCTGGGATTCAAAGGAGCCGCTGTCATGCTGTCGCGGAGATGATCCGCAAT 547  
 QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAspGlnGlu 29  
 DB 546 GCCACGACCGCGCGCATCTCTCCAGGCTCCCGGCAACTACAGACACACTCACC 487  
 QY 30 ArgArgTyrPheThrAlaProProGlyTyrArg 41  
 DB 486 TGTCACTGGCTGCTGAGCTCTGAGGCGCACGG 451

587/c  
 ABK43587 standard; cDNA: 810 BP.

ABK43587;  
 05-JUN-2002 (first entry)

DNA encoding novel central nervous system protein #167.

Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;  
 nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 adenocarcinoma; reproductive system disorder; testicular feminisation;  
 endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 respiratory disorder; renal disorder; kidney failure; blood disorder;  
 myocardial infarction; wound healing; cell proliferation; skin aging;  
 food additive; food preservative; gene therapy; gene; ss.

OS Homo sapiens.

PN WO20015318-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01332.

31-JAN-2000; 2000US-0179065.  
 24-FEB-2000; 2000US-0180628.  
 24-FEB-2000; 2000US-0184664.  
 02-MAR-2000; 2000US-0186350.  
 16-MAR-2000; 2000US-0189874.  
 17-MAR-2000; 2000US-0190076.  
 18-APR-2000; 2000US-0198123.  
 19-MAY-2000; 2000US-0205515.  
 07-JUN-2000; 2000US-0209467.  
 28-JUN-2000; 2000US-0214886.  
 30-JUN-2000; 2000US-0215135.  
 07-JUL-2000; 2000US-0216647.  
 07-JUL-2000; 2000US-0216880.  
 11-JUL-2000; 2000US-0217487.  
 14-JUL-2000; 2000US-0217496.  
 26-JUL-2000; 2000US-0218290.  
 26-JUL-2000; 2000US-0220963.  
 26-JUL-2000; 2000US-0220964.  
 14-AUG-2000; 2000US-0224518.  
 14-AUG-2000; 2000US-0224519.  
 14-AUG-2000; 2000US-0225213.  
 14-AUG-2000; 2000US-0225214.  
 14-AUG-2000; 2000US-0225266.  
 14-AUG-2000; 2000US-0225267.  
 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225457.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.



```

CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX Sequence 363 BP; 95 A; 95 C; 91 G; 81 T; 1 other.
S0
Alignment Scores:
Pred. No.: 0.0188 Length: 363
Score: 93.00 Matches: 14
Percent Similarity: 67.74% Conservative: 7
Best Local Similarity: 45.16% Mismatches: 10
Query Match: 39.08% Indels: 0
DB: 21 Gaps: 0
US-09-874-198-1 (1-41) x AAC01735 (1-363)
OY 11 ValbheglargleualaaserProglypheProglyglutylalaaaspsclnluarg 30
Db 183 ATGTTGGCCGAGATCCAGTCGCTGGTTATCCAGACTCCATCCAGTCAGATGACAGGNG 242
31 ArgTtrphtleuthralaProProglylTyArg 41
243 ACTTGAATATCACTGTCGCCAGATGGGTTTCGG 275
RESULT 7
AAC00227
ID AAC00227 standard; cDNA: 397 BP.
XX
AC AAC00227:
XX
XX 06-OCT-2000 (first entry)
DE Human secreted protein 5' EST, SEQ ID NO: 225.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 9905-0122487.
XX
PA (GEST ) GENSET.
XX
Dumas Milne Edwards J, Duclert A, Giordano J;
WPI; 2000-500381/45.
DR P-PSDB; AAG00221.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and Chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 225; 71pp + CD-ROM; English.
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.

```

XX	SO	Sequence	397 BP; 100 A; 104 C; 99 G; 93 T; 1 other;
XX		Alignment Scores:	
XX		Pred. No.:	0.0208
XX		Percent Similarity:	93.00
XX		Best Local Similarity:	67.74%
XX		Query Match:	45.16%
XX			39.08%
XX			21
XX		Gaps:	0
XX			
XX	OY	US-09-874-198-1 (1-41) x AAC00227 (1-397)	
XX	Db	11 ValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspIleuLary 30	
XX		:::     ::: ::   :::     :::	
XX	Db	184 ATGTTCGCCAGATCCGCCTGGTTATCCAGACTCCTCATCCAGTATTCACAGAGTG 243	
XX	OY	31 ArgTrpThrLeuThrAlaProPheGlyTyrArg 41	
XX	Db	244 ACTTGGAATATCACTGTCCTCCAGATGGGTTTCGG 276	
XX		:::         :::	
XX	RESULT 8		
XX	ID	AHA43512 standard; cDNA; 3863 BP.	
XX	XX	AHA43512:	
XX	DT	13-DEC-2001 (first entry)	
XX	DE	cDNA encoding protease PRTS-1..	
XX	KW	Human; protease; PRTS; gastrointestinal disorder; peptic esophagitis;	
XX	KM	indigestion; gastritis; cardiovascular disorder; reproductive disorder;	
XX	KW	hyperensive heart disease; myocardial infarction; autoimmune disorder;	
XX	KM	inflammatory disorder; rheumatoid arthritis; cell proliferative disorder;	
XX	KW	arteriosclerosis; cancer; epithelial disorder; eczema; endometriosis;	
XX	KM	neurological disorder; Huntington's disease; ss.	
XX	OS	Homo sapiens.	
XX	XX		
XX	FH	Location/Qualifiers	
XX	FT	CDS	/tag= "a"
XX	FT		/product= "PRTS-1"
XX	XX		
XX	PN	WO200171004-A2.	
XX	PD	27-SEP-2001.	
XX	PF	16-MAR-2001; 2001MO-US08441.	
XX	PR	17-MAR-2000; 2000US-190708P.	
XX	PR	30-MAR-2000; 2000US-193182P.	
XX	PR	14-APR-2000; 2000US-197086P.	
XX	PR	20-APR-2000; 2000US-199022P.	
XX	PR	28-APR-2000; 2000US-200227P.	
XX	PA	(INCY-) INCYTE GENOMICS INC.	
XX	PI	Yue H, Lu DAM, Policky JL, Delegeane AM, Tribouley CM, Khan FA;	
XX	PT	An-Young J, Bandman O, Lal P, Borowsky ML, Gandhi AR, Hillman JL;	
XX	PI	Tang YT, Burford N, Baughn MR, Nguyen DB, Yao MG, Walla NK, He A;	
XX	PI	Hatalia A, Lu Y, Patterson C;	
XX	DR	WPI: 2001-611509/70.	
XX	PT	p-PDB: AAB47559.	
XX	XX	New polypeptide for treating gastrointestinal, cardiovascular and	
XX	PT	autoimmune disorders, comprises novel human proteases (PRTS) and	
XX	PT	polynucleotides -	
XX	XX		
XX	SS	Claim 5; Page 122-23; 129pp; English.	
XX	XX		



CC The sequences given in AHA3512-22 encode novel human proteases PRS-1  
CC to -11. These proteases can be administered in pharmaceutical compounds  
CC for the treatment of diseases or conditions associated with decreased  
CC expression of functional protease. Disorders which can be treated  
CC include gastrointestinal (peptic esophagitis, indigestion and  
CC gastritis), cardiovascular (hypertensive heart disease, myocardial  
CC infarction), autoimmune/inflammatory (rheumatoid arthritis), cell  
CC proliferative (arteriosclerosis, cancers), epithelial (eczema),  
CC neurological (Huntington's disease), and reproductive (endometriosis)  
CC disorders. The protein encoded by this sequence shows homology  
CC with the human precursor of P100 serine protease of Ra-reactive factor.  
XX  
SQ Sequence 3863 BP; 923 A; 1063 C; 991 G; 886 T; 0 other:

## Alignment Scores:

Pred. No.: 0.264 Length: 3863  
Score: 93.00 Matches: 14  
Percent Similarity: 67.74% Conservative: 7  
Local Similarity: 45.16% Mismatches: 10  
Match: 39.08% Indels: 0  
Gaps: 22

US-09-874-198-1 (1-41) x AHA3512 (1-3863)

QY 11 ValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArg 30  
:::|||||::: |||||||::: ||| :||| |||  
DB 164 ATGTTGGCCAGATCCAGTCGCTGCTATCCAGACTCCCTATCCAGTATTCAGAGGTG 223  
CC

QY 31 ArgTfThleuThralaProProGlyTyrArg 41  
||| :||| ||| |||:::||||  
DB 224 ACTTGAATATCTCTCCACATGGGTTTCGG 256  
XX

## RESULT 9

AAF83985

ID AAF83985 standard; DNA; 3895 BP.

AC AAF83985;

DT 06-AUG-2001 (first entry)

XX Human serine protease MASP-3 encoding DNA.  
XX  
XX Mannan-binding lectin associated serine protease-3; MASP-3; MBL; human;  
KW mannan-binding lectin; complement; infection; cancer; cytostatic;  
KW immunomodulator; neuroprotective; anticonvulsive; antirheumatic;  
KW antiafleritic; antianemic; antiinflammatory; antipsoriatic; antidiabetic;  
KW antiatherosclerotic; ds.  
XX

Homo sapiens.

Key Location/Qualifiers  
FT CDS 91..2277  
FT /tag=a  
FT /product="MASP-3"  
XX

WO200140451-A2.

07-JUN-2001.

30-NOV-2000; 2000WO-DK00659.

02-DEC-1999; 99DK-0001721.

21-JUL-2000; 2000DK-0001126.

(JENS/) JENSENIUS J C.

(THIE/) THIEL S.

Jensenius JC, Thiel S;

WPI: 2001-374820/39.

P-PSDB; AAB5060.

Novel pure mannan-binding lectin associated serine protease

PT polypeptides and polynucleotides encoding the protein, used for  
PT treating reoxygenated ischemic tissues, mannon-binding lectin  
PT deficiency, multiple sclerosis  
XX  
PS Claim 18; Page 88-96; 99pp: English.

XX The invention relates to a mannan-binding lectin (MBL) associated serine  
CC protease-3 (MASP-3) polypeptide. MASP-3 exerts an inhibitory effect on  
CC the complement activation particularly when bound to MBL/MASP-2 complexes  
CC and directly activates complement system through binding to MBL. The  
CC MASP-3 polypeptides, polynucleotides and modulators are useful for  
CC preparing a pharmaceutical composition for treating aberrant MASP-3  
CC activity such as infections, cancer, MBL-deficiency, disorders of the  
CC immune system and reproductive system. The MASP-3 polypeptide is also  
CC used for treating diseases associated with human immunodeficiency virus,  
CC multiple sclerosis, myasthenia gravis, epilepsy, rheumatoid arthritis,  
CC vasculitis, autoimmune hemolytic anemia, Crohn's disease, asthma,  
CC diabetes, psoriasis, multiple myeloma, atherosclerosis etc. It is also  
CC useful for inhibiting activation of C4 complement by inhibiting the MBL  
CC pathway, inhibiting MASP-2 activity, inhibiting or treating an  
CC inflammatory condition related to complement activation through MBL/MASP  
CC complexes. MASP-3 is also useful for treating an inflammatory condition  
CC resulting from an autoimmune condition after acute myocardial infarction  
CC or brain ischemia. It is also useful for treating an individual suffering  
CC from a disorder resulting from an imbalanced cytokine network. The  
CC present sequence represents a DNA encoding a human MASP-3 polypeptide.

SQ Sequence 3895 BP; 952 A; 1068 C; 987 G; 888 T; 0 other:

## Alignment Scores:

Pred. No.: 0.267 Length: 3895  
Score: 93.00 Matches: 14  
Percent Similarity: 67.74% Conservative: 7  
Best Local Similarity: 45.16% Mismatches: 10  
Query Match: 39.08% Indels: 0  
DB: 22 Gaps: 0

US-09-874-198-1 (1-41) x AAF83985 (1-3895)

QY 11 ValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArg 30  
:::|||||::: |||||||::: ||| :||| |||  
DB 169 ATGTTGGCCAGATCCAGTCGCTGCTATCCAGACTCCCTATCCAGTATTCAGAGGTG 228  
CC

QY 31 ArgTfThleuThralaProProGlyTyrArg 41  
||| :||| ||| |||:::||||  
DB 229 ACTTGAATATCTCTCCACATGGGTTTCGG 261  
XX

## RESULT 10

AAF74442

ID AAF74442 standard; cDNA; 1988 BP.

AC AAF74442;

DT 09-MAY-2001 (first entry)

DE Human PRO11 nucleotide sequence SEQ ID NO:21.

XX Human; PRO; PROX; cytostatic; immunomodulatory; reproduction;  
KW gene therapy; cell proliferation; differentiation disorder; cancer;  
KW immune associated disorder; gestational disease; pre-clampsia; ss.  
XX

Homo sapiens.

WO200110902-A2.

15-FEB-2001.

11-AUG-2000; 2000WO-US21857.

11-AUG-1999; 99US-0148433.

10-AUG-2000; 2000US-0148433.

(CURA-) CURAGEN CORP.

```
XX Shinkets RA, Fernandes E;
XX
XX WPI: 2001-147509/15.
DR P-PSDB: AAB70541.
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
XX
PS Claim 8: Page 38-40; 166pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
CC X is an integer from 1 to 17). PROX polypeptides have cytostatic,
CC immunomodulatory and reproduction activities, and can be used in gene
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
CC nucleic acids and antibodies are useful in the manufacture of a
CC medicament for treating a syndrome associated with a PROX-associated
CC disorder, e.g. a cell proliferation and/or differentiation disorder
CC (e.g. cancer or immune associated disorders) and a gestational disease
CC (e.g. pre-clampsia). They are also used for screening for a modulator of
CC activity or of latency or predisposition to a PROX-associated disorder.
CC AAF74432 to AAF74448 encode the specifically claimed human PROX
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
SQ Sequence 1988 BP; 419 A; 695 C; 512 G; 362 T; 0 other:

Alignment Scores:
Pred. No.: 0.189 Length: 1988
Score: 91.50 Matches: 21
Percent Similarity: 46.15% Conservative: 3
Best Local Similarity: 40.38% Mismatches: 13
Query Match: 38.45% Indels: 15
DB: 22 Gaps: 2

US-09-874-198-1 (1-41) x AAF74442 (1-1988)

OY 5 ProlystTTP-----ProGluProVal----- 11
DB 1384 CCCTTCTGGGATTCAAGAGCCGCTGTCATCGCTTGGCGGCGAGTGATCCGCAAT 1443
OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29
DB 1444 GGCACACCGCGCGATGCTCTCTCCAGGCTTCCGCGCACTACAGCAACACTCCACC 1503
OY 30 ArgArgTTrpThrLeuThrAlaProProGlyTyrArg 41
DB 1504 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 1539
AAAF74440
ID AAF74440 standard; cDNA; 2127 BP.
XX
XX AAF74440;
AC
XX
XX 09-MAY-2001 (first entry)
DT
XX
XX Human PRO9 nucleotide sequence SEQ ID NO:17.
DE
XX
XX Human; PRO; PROX; cytostatic; immunomodulatory; reproduction;
KW gene therapy; cell proliferation; differentiation disorder; cancer;
KW immune associated disorder; gestational disease; pre-clampsia; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200110902-A2.
PN
XX
XX 15-FEB-2001.
PD
XX
XX 11-AUG-2000; 2000MO-US21857.
PF
XX
XX 11-AUG-1999; 99US-0148433.
PR
```

```
PR 10-AUG-2000; 2000US-0148433.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Fernandes E;
XX
XX WPI: 2001-147509/15.
DR P-PSDB: AAB70539.
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
XX
PS Claim 8: Page 33-35; 166pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
CC X is an integer from 1 to 17). PROX polypeptides have cytostatic,
CC immunomodulatory and reproduction activities, and can be used in gene
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
CC nucleic acids and antibodies are useful in the manufacture of a
CC medicament for treating a syndrome associated with a PROX-associated
CC disorder, e.g. a cell proliferation and/or differentiation disorder
CC (e.g. cancer or immune associated disorders) and a gestational disease
CC (e.g. pre-clampsia). They are also used for screening for a modulator of
CC activity or of latency or predisposition to a PROX-associated disorder.
CC AAF74432 to AAF74448 encode the specifically claimed human PROX
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
SQ Sequence 2127 BP; 430 A; 764 C; 557 G; 375 T; 1 other:

Alignment Scores:
Pred. No.: 0.204 Length: 2127
Score: 91.50 Matches: 21
Percent Similarity: 46.15% Conservative: 3
Best Local Similarity: 40.38% Mismatches: 13
Query Match: 38.45% Indels: 15
DB: 22 Gaps: 2

US-09-874-198-1 (1-41) x AAF74440 (1-2127)

OY 5 ProlystTTP-----ProGluProVal----- 11
DB 1384 CCCTTCTGGGATTCAAGAGCCGCTGTCATCGCTTGGCGGCGAGTGATCCGCAAT 1443
OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29
DB 1444 GGCACACCGCGCGATGCTCTCTCCAGGCTTCCGCGCACTACAGCAACACTCCACC 1503
OY 30 ArgArgTTrpThrLeuThrAlaProProGlyTyrArg 41
DB 1504 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 1539
AAAF74441
ID AAF74441 standard; cDNA; 2127 BP.
XX
XX AAF74441;
AC
XX
XX 09-MAY-2001 (first entry)
DT
XX
XX Human PRO10 nucleotide sequence SEQ ID NO:19.
DE
XX
XX Human; PRO; PROX; cytostatic; immunomodulatory; reproduction;
KW gene therapy; cell proliferation; differentiation disorder; cancer;
KW immune associated disorder; gestational disease; pre-clampsia; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200110902-A2.
PN
XX
XX 15-FEB-2001.
PD
XX
```

PF 11-AUG-2000: 2000MO-US21857.  
XX  
XX 11-AUG-1999: 99US-0148433.  
PR 10-AUG-2000: 2000US-0148433.  
XX  
PA (CURA-) CURAGEN CORP.  
PI Shimkets RA, Fernandes E;  
XX  
DR WPI: 2001-147509/15.  
DR P-PSDB: AAB70540.  
XX  
PT Nucleic acids encoding secreted polypeptides, designated PROX  
PT polypeptides, useful for treating a syndrome associated with a  
PT PROX-associated disorder, e.g. cancer -  
XX  
PS  
XX  
XX Claim 8: Page 35-37; 166pp; English.  
XX  
XX The present invention describes isolated nucleic acids encoding secreted  
XX polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where  
XX X is an integer from 1 to 17). PROX polypeptides have cytostatic,  
XX immunomodulatory and reproduction activities, and can be used in gene  
XX therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,  
XX nucleic acids and antibodies are useful in the manufacture of a  
XX medicament for treating a syndrome associated with a PROX-associated  
XX disorder, e.g. a cell proliferation and/or differentiation disorder  
XX (e.g. cancer or immune associated disorders) and a gestational disease  
XX (e.g. pre-clampsia). They are also used for screening for a modulator of  
XX CC activity or of latency or predisposition to a PROX-associated disorder.  
XX CC AAF74432 to AAF74448 encode the specifically claimed human PROX  
XX polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.  
XX  
SQ Sequence 2127 BP; 431 A; 758 C; 557 G; 380 T; 1 other;  
XX  
XX Alignment Scores:  
Pred. No.: 0.204 Length: 2127  
Score: 91.50 Matches: 21  
Percent Similarity: 46.15% Conservative: 3  
Best Local Similarity: 40.38% Mismatches: 13  
Query Match: 38.45% Indels: 15  
Gaps: 2  
XX  
US-09-874-198-1 (1-41) x AAF74441 (1-2127)  
XX  
OY 5 ProlystTrp-----ProgluProval----- 11  
DB 1384 CCCTTCGGGATTCAAAGAGAGCCCGCTGCATCGCTCGCGGAGATGCCGCAAT 1443  
12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLutYrAlaAsnAspGlnGlu 29  
1444 GGCACACACCGCGCGATCGTCTCCAGGCTTCCCGGCAACACACACCTCACC 1503  
OY 30 ArgArgTrpThrLeuThrAlaProProGlyTYrArg 41  
DB 1504 TGTCACTGGCTGCTTGAAGCTCCTGAGGCGCACGCG 1539  
XX  
XX RESULT 13  
XX AAF74443  
XX ID AAF74443 standard; cDNA; 2143 BP.  
XX  
XX AAF74443;  
XX  
XX 09-MAY-2001 (first entry)  
XX  
XX Human PRO12 nucleotide sequence SEQ ID NO:23.  
XX  
XX Human: PRO; PROX; cytostatic; immunomodulatory; reproduction;  
XX gene therapy; cell proliferation; differentiation disorder; cancer;  
XX immune associated disorder; gestational disease; pre-clampsia; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200110902-A2.  
XX  
XX PN

XX  
XX 15-FEB-2001.  
PD  
XX  
XX 11-AUG-2000: 2000MO-US21857.  
PF  
XX  
XX 11-AUG-1999: 99US-0148433.  
PR 10-AUG-2000: 2000US-0148433.  
XX  
XX (CURA-) CURAGEN CORP.  
PA Shimkets RA, Fernandes E;  
XX  
DR WPI: 2001-147509/15.  
DR P-PSDB: AAB70542.  
XX  
PT Nucleic acids encoding secreted polypeptides, designated PROX  
PT polypeptides, useful for treating a syndrome associated with a  
PT PROX-associated disorder, e.g. cancer -  
XX  
PS  
XX  
XX Claim 8: Page 41-43; 166pp; English.  
XX  
XX The present invention describes isolated nucleic acids encoding secreted  
XX polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where  
XX X is an integer from 1 to 17). PROX polypeptides have cytostatic,  
XX immunomodulatory and reproduction activities, and can be used in gene  
XX therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,  
XX nucleic acids and antibodies are useful in the manufacture of a  
XX medicament for treating a syndrome associated with a PROX-associated  
XX disorder, e.g. a cell proliferation and/or differentiation disorder  
XX (e.g. cancer or immune associated disorders) and a gestational disease  
XX (e.g. pre-clampsia). They are also used for screening for a modulator of  
XX CC activity or of latency or predisposition to a PROX-associated disorder.  
XX CC AAF74432 to AAF74448 encode the specifically claimed human PROX  
XX polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.  
XX  
SQ Sequence 2143 BP; 431 A; 769 C; 559 G; 383 T; 1 other;  
XX  
XX Alignment Scores:  
Pred. No.: 0.206 Length: 2143  
Score: 91.50 Matches: 21  
Percent Similarity: 46.15% Conservative: 3  
Best Local Similarity: 40.38% Mismatches: 13  
Query Match: 38.45% Indels: 15  
Gaps: 2  
XX  
US-09-874-198-1 (1-41) x AAF74443 (1-2143)  
XX  
OY 5 ProlystTrp-----ProgluProval----- 11  
DB 1384 CCCTTCGGGATTCAAAGAGAGCCCGCTGCATCGCTTCCGCGGAGATGCCGCAAT 1443  
12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLutYrAlaAsnAspGlnGlu 29  
1444 GGCACACACCGCGCGATCGTCTCCAGGCTTCCCGGCAACACACACCTCACC 1503  
OY 30 ArgArgTrpThrLeuThrAlaProProGlyTYrArg 41  
DB 1504 TGTCACTGGCTGCTTGAAGCTCCTGAGGCGCACGCG 1539  
XX  
XX RESULT 14  
XX AAF74438  
XX ID AAF74438 standard; cDNA; 3863 BP.  
XX  
XX AAF74438;  
XX  
XX 09-MAY-2001 (first entry)  
XX  
XX Human PRO7 nucleotide sequence SEQ ID NO:13.  
XX  
XX Human: PRO; PROX; cytostatic; immunomodulatory; reproduction;  
XX gene therapy; cell proliferation; differentiation disorder; cancer;  
XX immune associated disorder; gestational disease; pre-clampsia; ss.  
XX  
XX  
XX

OS Homo sapiens.  
XX  
PN WO200110902-A2.  
XX  
PD 15-FEB-2001.  
XX  
PF 11-AUG-2000; 2000WO-US21857.  
XX  
PR 11-AUG-1999; 99US-0148433.  
PR 10-AUG-2000; 2000US-0148433.  
XX  
PA (CURA-) CURAGEN CORP.  
PI Shimkets RA, Fernandes E;  
XX  
DR WPI: 2001-147509/15.  
DR P-PSDB; AAB70537.  
XX  
XX Nucleic acids encoding secreted polypeptides, designated PROX polypeptides, useful for treating a syndrome associated with a PROX-associated disorder, e.g. cancer -

Claim 8; Page 24-27; 16pp; English.

XX The present invention describes isolated nucleic acids encoding secreted  
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where  
CC X is an integer from 1 to 17). PROX polypeptides have cytostatic,  
CC immunomodulatory and reproduction activities, and can be used in gene  
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,  
CC nucleic acids and antibodies are useful in the manufacture of a  
CC medicament for treating a syndrome associated with a PROX-associated  
CC disorder, e.g. a cell proliferation and/or differentiation disorder  
CC (e.g. cancer or immune associated disorders) and a gestational disease  
CC (e.g. pre-clampsia). They are also used for screening for a modulator of  
CC activity or of latency or predisposition to a PROX-associated disorder.  
CC AAF74432 to AAF74448 encode the specifically claimed human PROX  
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.  
XX

SQ Sequence 3863 BP; 781 A; 1306 C; 1028 G; 747 T; 1 other;

Alignment Scores:

Pred. No.:	0.398	Length:	3863
Score:	91.50	Matches:	21
Percent Similarity:	46.15%	Conservative:	3
Best Local Similarity:	40.38%	Mismatches:	13
Query Match:	38.45%	Indels:	15
DB:	22	Gaps:	2

-874-198-1 (1-41) x AAF74438 (1-3863)

5 ProlystR-----ProgluProval----- 11  
DB 1384 CCCATCTGGGATTCMAAGGAGCCGTRATGCATGCTGTCGGCGGAGTGATCCGCAT 1443  
OY 12 -----PhelYArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29  
DB 1444 GCCACACCGCGCGCATGCTCTCCAGGCTTCCCGGCAACTACAGCAACACCTCACC 1503  
OY 30 ArgArgTyrThrLeuThrAlaProProGlyTyrArg 41  
DB 1504 TGTCACTGGCTGCTTGAAGGCTCTGAGGGCCAGCGG 1539

RESULT 15

AAE74439

ID AAE74439 standard; cDNA; 3879 BP.

XX AAE74439;

XX 09-MAY-2001 (first entry)

DE Human PRO8 nucleotide sequence SEQ ID NO:15.

XX Human; PRO; PROX; cytostatic; immunomodulatory; reproduction;

KW gene therapy; cell proliferation; differentiation disorder; cancer;  
KW immune associated disorder; gestational disease; pre-clampsia; ss.

OS Homo sapiens.

XX WO200110902-A2.

XX 15-FEB-2001.

XX 11-AUG-2000; 2000WO-US21857.

XX 11-AUG-1999; 99US-0148433.

XX 10-AUG-2000; 2000US-0148433.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Fernandes E;

XX WPI: 2001-147509/15.

XX P-PSDB; AAB70538.

XX Nucleic acids encoding secreted polypeptides, designated PROX polypeptides, useful for treating a syndrome associated with a PROX-associated disorder, e.g. cancer -

Claim 8; Page 28-31; 16pp; English.

XX The present invention describes isolated nucleic acids encoding secreted  
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where  
CC X is an integer from 1 to 17). PROX polypeptides have cytostatic,  
CC immunomodulatory and reproduction activities, and can be used in gene  
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,  
CC nucleic acids and antibodies are useful in the manufacture of a  
CC medicament for treating a syndrome associated with a PROX-associated  
CC disorder, e.g. a cell proliferation and/or differentiation disorder  
CC (e.g. cancer or immune associated disorders) and a gestational disease  
CC (e.g. pre-clampsia). They are also used for screening for a modulator of  
CC activity or of latency or predisposition to a PROX-associated disorder.  
CC AAF74432 to AAF74448 encode the specifically claimed human PROX  
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.  
XX

SQ Sequence 3879 BP; 782 A; 1311 C; 1030 G; 755 T; 1 other;

Alignment Scores:

Pred. No.:	0.4	Length:	3879
Score:	91.50	Matches:	21
Percent Similarity:	46.15%	Conservative:	3
Best Local Similarity:	40.38%	Mismatches:	13
Query Match:	38.45%	Indels:	15
DB:	22	Gaps:	2

US-09-874-198-1 (1-41) x AAE74439 (1-3879)

5 ProlystR-----ProgluProval----- 11  
DB 1384 CCCATCTGGGATTCMAAGGAGCCGTRATGCATGCTGTCGGCGGAGTGATCCGCAT 1443  
OY 12 -----PhelYArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29  
DB 1444 GCCACACCGCGCGCATGCTCTCCAGGCTTCCCGGCAACTACAGCAACACCTCACC 1503  
OY 30 ArgArgTyrThrLeuThrAlaProProGlyTyrArg 41  
DB 1504 TGTCACTGGCTGCTTGAAGGCTCTGAGGGCCAGCGG 1539

RESULT 16

ABK28635

ID ABK28635 standard; cDNA; 4034 BP.

XX ABK28635;

XX 09-APR-2002 (first entry)

DE Human cDNA encoding secreted protein SECP2.  
 XX  
 XX Human; ss: gene: SECP: antiinflammatory; cytosolic; cardiant.  
 KW immunosuppressive; antiviral; anti-HIV; antiallergic; antineoplastic;  
 KW muscular active general; anticonvulsant; nootropic; neuroprotective;  
 KW antiallergic; hypotensive; cardiovascular disorder; atherosclerosis;  
 KW hypertension; myocardial infarction; autoimmune disorder;  
 KW inflammatory disorder; AIDS; acquired immunodeficiency syndrome;  
 KW allergy; rheumatoid arthritis; cell proliferative disorder; cancer;  
 KW developmental disorder; Duchenne muscular dystrophy;  
 KW neurological disorder; epilepsy; Alzheimer's disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX MO200198353-A2.  
 PN  
 XX  
 XX 27-DEC-2001.  
 PD  
 XX  
 XX 20-JUN-2001; 2001WO-US19862.  
 XX  
 XX 20-JUN-2000; 2000US-212890P.  
 XX 23-JUN-2000; 2000US-213466P.  
 PR 27-JUN-2000; 2000US-214601P.  
 PR 31-JUL-2000; 2000US-222372P.  
 PR 08-SEP-2000; 2000US-231435P.  
 PR 15-SEP-2000; 2000US-232889P.  
 XX  
 XX (INCYTE GENOMICS INC.  
 PA  
 XX Hillman JL, Tang YT, Yue H, Elliott VS, Tribouley CM, Lee EA;  
 PI Ramkumar J, Lal P, Xu Y, Warren BA, Hafalia AJA, Baughn MR;  
 PI Azimzal Y, Batera S, Burford N, Yao MG, Nguyen DB, Lu DAM;  
 PI Wella NK, Gandhi AR, Au-Young J, Patterson C;  
 XX  
 XX WPI: 2002-090431/12.  
 DR P-PSDB: AA081976.  
 XX  
 XX Forty four human secreted proteins (referred to as SECP-1 to SECP-44),  
 PT useful in the diagnosis, treatment and prevention of cardiovascular  
 PT (e.g. atherosclerosis), autoimmune/inflammatory (e.g. allergies) and  
 PT cell proliferative disorders -  
 XX  
 XX Claim 5; Page 166-167; 195pp; English.  
 XX  
 CC The invention relates to forty four human secreted proteins (referred to  
 CC as SECP-1 to SECP-44) and the nucleic acids encoding them. Also  
 CC included are a host cell transformed with the nucleic acid, a  
 CC transgenic animal comprising the nucleic acid, an anti-SECP  
 CC antibody, use of the SECP proteins in isolating agonists and antagonists  
 CC of SECP activity and a method of isolating compounds which alter the  
 CC expression of the SECP nucleic acid. The SECP polynucleotides and  
 CC polypeptides are useful in the diagnosis, treatment and prevention of  
 CC cardiovascular (e.g. atherosclerosis, hypertension, myocardial  
 CC infarction), autoimmune/inflammatory (e.g. acquired immunodeficiency  
 CC syndrome (AIDS), allergies, rheumatoid arthritis), cell proliferative  
 CC (e.g. cancer), developmental (e.g. Duchenne and Becker muscular  
 CC dystrophy), and neurological (e.g. epilepsy, Alzheimer's disease)  
 CC disorders. Numerous other examples of each disorder are given in the  
 CC specification. The present sequence is a cDNA encoding a SECP protein.  
 CC  
 XX  
 XX Sequence 4034 BP; 840 A; 1305 C; 1062 G; 827 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. NO.: 0.417 Length: 4034  
 Score: 91.50 Matches: 21  
 Percent Similarity: 46.15% Conservative: 13  
 Best Local Similarity: 40.38% Mismatches: 15  
 Query Match: 38.45% Indels: 13  
 DB: 24 Gaps: 2  
 US-09-874-198-1 (1-41) x ABK28635 (1-4034)  
 OY 5 ProllyStrp-----ProgiuProval----- 11

Db 1217 CCCTCTGGGATTCAAAGAGACCCGCTGCTGCGGCGGAGATCCGCAMF 1276  
 OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29  
 Db 1277 GCCACGACCGCGCGCATGCTCTCCAGGCTCCCGGCGCACTACAGCAACACTCCACC 1336  
 OY 30 ArgArgTrpThrLeuThrAlaProProGlyTyrArg 41  
 Db 1337 TGTCACTGGCTGCTGAGGCTCCTGAGAGCGCAGCGG 1372  
 RESULT 17  
 AAD25344  
 ID AAD25344 strand; DNA: 4198 BP.  
 AC  
 XX AAD25344;  
 XX  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX  
 XX Human SEZ6 DNA.  
 DE  
 XX  
 XX Human; SEZ6; neural regeneration; seizure; infertility; gene therapy;  
 KW stroke; Alzheimer's disease; Huntington's disease; myasthenia gravis;  
 KW nootropic; neuroprotective; anticonvulsant; cerebroprotective; stroke;  
 KW vasotropic; neuronal growth; growth factor-mediated chemotaxis; trauma;  
 KW neurological disorder; muscular dystrophy; muscle injury; vulnery;  
 KW amyotrophic lateral sclerosis; multiple sclerosis; ischaemia; diabetes;  
 KW epilepsy; Parkinson's disease; sexual development; gene mapping;  
 KW impotence; libido; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 191..2732  
 FT /\*tag= a  
 FT /product= "Human SEZ6 protein"  
 FT sig-peptide 191..262  
 FT /\*tag= b  
 FT mat-peptide 263..2749  
 FT /\*tag= c  
 FT /product= "Mature hSEZ6 protein"  
 FT misc-feature 191..2749  
 FT /\*tag= d  
 FT /note= "this region is specifically claimed as SEQ  
 FT ID NO: 2 in claim 13 of the specification"  
 XX  
 XX WO200183552-A2.  
 XX  
 XX 08-NOV-2001.  
 XX  
 XX 17-APR-2001; 2001WO-US10809.  
 XX  
 XX 28-APR-2000; 2000US-200200P.  
 XX  
 XX (ELIL ) LILLY & CO ELI.  
 PA  
 XX  
 XX Su EW;  
 PI  
 XX  
 XX WPI: 2002-082840/11.  
 DR P-PSDB: AAE15853, AAE15854.  
 XX  
 XX Novel human SEZ6 polypeptide useful for inducing neural regeneration,  
 PT inhibiting neural degeneration, preventing seizures, and for treating  
 PT infertility, Alzheimer's disease, stroke, Huntington's  
 PT disease -  
 XX  
 XX Claim 13; Page 117-118; 127pp; English.  
 CC The invention relates to an isolated human SEZ6 (hSEZ6) polypeptide and  
 CC its corresponding DNA molecule. SEZ6 is useful for enhancing neuronal  
 CC growth, neurite outgrowth, neuronal regeneration, neuronal survival,  
 CC growth factor-mediated chemotaxis, altered behavioural patterns, e.g.,  
 CC sleep or eating disorders and for treating neurological disorders such

CC as trigeminal neuralgia, Bell's palsy, myasthenia gravis, muscular  
CC dystrophy, muscle injury, invertebrate disk syndrome, thoracic outlet  
CC destruction syndrome, amniotic lateral sclerosis, multiple sclerosis,  
CC ischaemia associated with stroke, neuropathy associated with diabetes,  
CC spinal cord trauma, facial nerve crush and other trauma, Huntington's  
CC disease. SE26 is useful for treating a patient suffering from a  
CC neurological disorder such as epilepsy, Alzheimer's disease, Parkinson's  
CC disease, seizure related disorder or a disorder associated with stroke.  
CC SE26 DNA is useful as probes for gene mapping and for detecting  
CC transcription, translation and/or expression of hSE26 polypeptide in  
CC human tissue. A transgenic animal is useful as an animal model in  
CC research and drug development procedures, and for testing compounds or  
CC other treatment modalities which may prevent, suppress or cure a  
CC pathology or disease associated with hSE26 activities. SE26 is also  
CC useful for treating abnormal primary or secondary sexual development,  
CC e.g., impotence, infertility or reduced libido. The hSE26 polynucleotide  
CC is useful for treating the above mentioned disorders by gene therapy  
CC techniques. The present sequence is human SE26 DNA.

CC Sequence 4198 BP; 854 A; 1386 C; 1119 G; 839 T; 0 other;

CC Alignment Scores:

Seq. No.:	0.436	Length:	4198
Score:	91.50	Matches:	21
Percent Similarity:	46.15%	Conservative:	13
Best Local Similarity:	40.38%	Mismatches:	15
Query Match:	38.45%	Indels:	2
DB:	24	Gaps:	2

US-09-874-198-1 (1-41) x AAD25344 (1-4198)

QY 5 ProlystP-----ProGluProVal----- 11

Db 1397 CCCTTCGGGATTCGAAGGAGCCGCTGTCATCGCTCTCGCGGAGATCCGCAAT 1456

QY 12 -----PhcGlyArgLeuAlaSerProGlyPheProGlyGluTyralaasnaspGlnGlu 29

Db 1457 GCCACGACCGCGCGCATCGTCTCCAGGCTCCGCGGCACTACACACACCTCACC 1516

QY 30 ArgArgTyrPheThrAlaProProGlyTyrArg 41

Db 1517 TGTCACTGGCTGCTGAGGCTCCTGAGGCGCACGGG 1552

RESULT 18

ABL10283

ID ABL10283 standard; cDNA; 9455 BP.

XX ABL10283;

26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 25331.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

PT P-PSDB; ABB66180.

DR

XX

XX

XX

XX

XX

XX

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PS Claim 1: SEQ ID NO 25331; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC SQ Sequence 9455 BP; 2333 A; 2397 C; 2484 G; 2241 T; 0 other;

QY 13 GLYArgLeuAlaSerProGlyPheProGlyGluTyralaasnaspGlnGluArgTyrP 32

Db 1738 GCGGTGCTGGATCTCCGGGCTACTCTGTCAGGCGGACCGAAGTCACTCTCCTGG 1797

QY 33 ThrLeuThrAlaProProGlyTyrArg 41

Db 1798 CACTGACGCGCGCGTTGGCTACCGG 1824

RESULT 19

ABL10282

ID ABL10282 standard; cDNA; 13036 BP.

XX ABL10282;

26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 25328.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

PT P-PSDB; ABB66179.

DR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -

XX Claim 1: SEQ ID NO 25328; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB161757-AB161757).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SO Sequence 13036 BP; 3489 A; 3058 C; 3137 G; 3352 T; 0 other:  
  
Alignment Scores:  
NO.: 12 Length: 13036  
Matches: 84.00  
Conservative: 68.97%  
Local Similarity: 55.17%  
Mismatch: 35.29%  
Indels: 0  
Gaps: 0  
DB: 23  
  
US-09-874-198-1 (1-41) x ABL10282 (1-13036)  
QY 13 GlyArgLeuAlaSerProGlyPheProGlyGlyTyrAlaAsnAspGlnGluArgTyr 32  
DB 3161 GGGCTGCTGGATCTCCGGCTATCTCGTACAGCGGACCGAGACTGCTCGCTGG 3220  
QY 33 ThrLeuThrAlaProProGlyTyrArg 41  
DB 3221 CACTGACGGCGCGCTTGGCTACCG 3247  
  
RESULT 20  
ABK63805  
ID ABK63805 standard; cDNA; 4344 BP.  
XX  
AC ABK63805;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Rat sequence differentially expressed in response to a hepatotoxin #1712.  
XX  
KM Rat: ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
differential expression; centrilobular necrosis; steatosis.  
  
Rattus norvegicus.  
WO200210453-A2.  
XX  
PD 07-FEB-2002.  
XX  
PF 30-JUL-2001; 2001WO-US23872.  
XX  
PR 31-JUL-2000; 2000US-222040P.  
XX  
PR 02-NOV-2000; 2000US-244880P.  
XX  
PR 11-MAY-2001; 2001US-290029P.  
XX  
PR 15-MAY-2001; 2001US-290645P.  
XX  
PR 22-MAY-2001; 2001US-292336P.  
XX  
PR 06-JUN-2001; 2001US-295798P.  
XX  
PR 13-JUN-2001; 2001US-297457P.  
XX  
PR 19-JUN-2001; 2001US-298884P.  
XX  
PR 09-JUL-2001; 2001US-303459P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Mendlick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;  
XX  
DR WPI; 2002-241625/29.  
XX  
PT Predicting toxic effects of compounds or the progression of these toxic

PT effects by determining the changes in gene expression in tissues or  
PT cells exposed to the toxin and comparing these to gene expression in  
PT unexposed tissues or cells -  
XX  
PS Claim 1: Seq ID No 1712; 239pp; English.  
XX  
XX  
CC The invention relates to methods for predicting toxic effects of  
CC compounds or the progression of these toxic effects by determining the  
CC global changes in gene expression in tissues or cells exposed to the  
CC toxin and comparing these to gene expression in unexposed tissues or  
CC cells. Also included are methods of predicting at least one toxic  
CC effect of a compound or progression of a toxic effect, preferably the  
CC hepatotoxicity of a compound, comprising detecting the level of  
CC expression in a tissue or cell sample exposed to the compound of two or  
CC more genes listed in the specification, where differential expression of  
CC the genes is indicative of at least one toxic effect or progression.  
CC The method can also be used to identify an agent which modulates the  
CC toxic response and predict cellular pathways that a compound modulates  
CC in a cell. The method utilizes a set of at least two probes (on a solid  
CC support in kit form), where each of the probes comprises a sequence that  
CC specifically hybridizes to a gene listed in the specification, a computer  
CC system comprising a database containing information identifying the  
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
CC set of genes comprising at least two genes listed in the specification,  
CC and a user interface to view the information used to present information  
CC identifying the expression level in a tissue or cell of at least one gene  
CC listed in the specification. The method is useful for elucidating global  
CC changes in gene expression and for identifying toxicity markers in  
CC tissues or cell exposed to a known toxin. The genes may be used as  
CC toxicity markers in drug screening and toxicity assays. The genes and  
CC gene expression information may be used as diagnostic markers for the  
CC prediction or identification of the physiological state of tissue or cell  
CC sample that has been exposed to a compound or agent. Hepatotoxicity  
CC is characterized by centrilobular necrosis and steatosis. The present  
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
CC which is differentially expressed in response to a hepatotoxic agent.  
XX  
SQ Sequence 4344 BP; 1035 A; 1099 C; 1103 G; 1107 T; 0 other:  
  
Alignment Scores:  
Pred. No.: 4.02 Length: 4344  
Score: 83.50  
Percent Similarity: 41.18%  
Best Local Similarity: 31.37%  
Query Match: 35.08%  
DB: 24 Indels: 13  
Gaps: 1  
  
US-09-874-198-1 (1-41) x ABK63805 (1-4344)  
QY 4 GlyProLystrProGluProValPhe----- 12  
DB 2566 GGTTCAGTTTGGACAACTGCTCTTGTAACTACTGTTGGAGGTTTCTGACTGCA 2625  
QY 13 -----GlyArgLeuAlaSerProGlyPheProGlyGlyTyrAlaAsnAspGlnGluArg 30  
DB 2626 CTCTCTGGCAATTTTCTACCCACTACTACCTGGAGCTATCTAATATGCCAGATGT 2685  
QY 31 ArgTrpThrLeuThrAlaProProGlyTyrArg 41  
DB 2686 TTGTGCAACATTGAGTCCCAACAACTACCGC 2718  
  
Search completed: January 11, 2003, 11:16:24  
Job time : 358 secs

**THIS PAGE BLANK (USPTO)**